PCT/US 97/03313

U.S. PATENT APPLICATION

SERIAL NUMBER	FILING DATE	CLASS	GROUP ART UNIT
60/013,706 PROVISIONAL	03/20/96		

APPLICANT

SANDRA L. VOLRATH, DURHAM, NC; MARIE A. FADELLA, RALEIGH, NC; SHARON L. POTTER, RALEIGH, NC; ERIC B. WARD, DURHAM, NC.

CONTINUING DATA********

VERIFIED

———

REC'D 15 APR 1997
WIPO PCT

FOREIGN/PCT APPLICATIONS********

VERIFIED

———

PRIORITY DOCUMENT

FOREIGN FILING LICENSE GRANTED 04/10/96

STATE OR COUNTRY	SHEETS DRAWING	TOTAL CLAIMS	INDEPENDENT CLAIMS	FILING FEE RECEIVED	ATTORNEY DOCKET NO.
NC	0			$150.00	PROV/CGC1847

ADDRESS

MICHAEL W GLYNN
CIBA GEIGY CORPORATION
520 WHITE PLAINS RD PATENT DEPT
P O BOX 2005
TARRYTOWN NY 10591-9005

TITLE

DNA MOLECULES ENCODING PLANT PROTOPORPHYRINOGEN OXIDASE AND INHIBITOR-RESISTANT MUTANTS THEREOF

This is to certify that annexed hereto is a true copy from the records of the United States Patent and Trademark Office of the application which is identified above.

By authority of the Commissioner of Patents and Trademarks

Date MAY 17 1997 Certifying Officer

PATENT APPLICATION SERIAL NO. 60/012705

U.S. DEPARTMENT OF COMMERCE
PATENT AND TRADEMARK OFFICE
FEE RECORD SHEET

PTO-1556
(5/87)

60/012705

Docket Number	PROV/CGC1847

CERTIFICATE OF MAILING BY "EXPRESS MAIL"

EG826034391US	February 28, 1996
"Express Mail" mailing number	Date of Deposit

I hereby certify that this paper (along with any paper referred to as being attached or enclosed) is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above and is addressed to the Commissioner of Patents and Trademarks, Washington, D.C. 20231

Priscilla Tarleton	Priscilla Tarleton
(Typed or printed name of person mailing application)	(Signature of person mailing application)

To: Box: Provisional Patent Applications
Assistant Commissioner of Patents
Washington, D.C. 20231

PROVISIONAL PATENT APPLICATION COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION for patent under 37 CFR 1.53 (b)(2).

INVENTOR(s)/APPLICANT(s)

LAST NAME	FIRST NAME	MIDDLE INITIAL	RESIDENCE (CITY AND EITHER STATE OR FOREIGN COUNTRY)
Volrath	Sandra	L.	Durham, North Carolina NC
Pacella	Maria	A.	Raleigh, North Carolina NC
Potter	Sharon	L.	Raleigh, North Carolina NC
Ward	Eric	S.	Durham, North Carolina NC

TITLE OF THE INVENTION (280 characters max)
DNA Molecules Encoding Plant Protoporphyrinogen Oxidase and Inhibitor-Resistant Mutants Thereof

CORRESPONDENCE ADDRESS

Michael W. Glynn
Ciba-Geigy Corporation
Patent Department

Mailing Address	State	Zip Code	Country
520 White Plains Road P.O. Box 2005 Tarrytown	NEW YORK	10591-9005	U.S.A.

ENCLOSED APPLICATION PARTS (check all that apply)

- [X] 85 pages of Specification (and any claims)
- [X] 1 page of Abstract (page 85)
- [ ] ___ sheets of Drawing(s)
- [ ] Other (specify)

METHOD OF PAYMENT

The Commissioner is hereby authorized to charge filing fees and credit fees required to Deposit Account Number 07-0570	PROVISIONAL FILING FEE AMOUNT ($) 150.00

[ ] U.S. Government agency and contract number _____ (if the invention was made by an agency of the United States Government or under a contract with an agency of the United States Government)

Respectfully submitted,

NAME James Scott Elmer
ATTORNEY/AGENT
REGISTRATION NO 36,129

Tel No (919) 541-8614
Date February 28, 1996

[ ] Additional inventors are being named on separately numbered sheets attached hereto

100 11/88

A Rev.

60/012705

DNA MOLECULES ENCODING PLANT PROTOPORPHYRINOGEN OXIDASE AND INHIBITOR-RESISTANT MUTANTS THEREOF

FIELD OF THE INVENTION

The invention relates generally to the plant enzyme protoporphyrinogen oxidase ("protox"). In particular, the invention relates to DNA molecules encoding this enzyme and to modified, inhibitor-resistant forms of this enzyme. The invention further relates to methods for tissue culture selection and herbicide application based on these modified forms.

BACKGROUND OF THE INVENTION

I. The Protox Enzyme and its Involvement in the Chlorophyll/Heme Biosynthetic Pathway

The biosynthetic pathways which lead to the production of chlorophyll and heme share a number of common steps. Chlorophyll is a light harvesting pigment present in all green photosynthetic organisms. Heme is a cofactor of hemoglobin, cytochromes, P450 mixed-function oxygenases, peroxidases, and catalases (*see, e.g.* Lehninger, Biochemistry, Worth Publishers, New York (1975)), and is therefore a necessary component for all aerobic organisms.

The last common step in chlorophyll and heme biosynthesis is the oxidation of protoporphyrinogen IX to protoporphyrin IX. Protoporphyrinogen oxidase (referred to herein as "protox") is the enzyme which catalyzes this last oxidation step (Matringe *et al., Biochem. J. 260:* 231 (1989)).

The protox enzyme has been purified either partially or completely from a number of organisms including the yeast *Saccharomyces cerevisiae* (Labbe-Bois and Labbe, In Biosynthesis of Heme and Chlorophyll, E.H. Dailey, ed. McGraw Hill: New York, pp. 235-285 (1990)), barley etioplasts (Jacobs and Jacobs, *Biochem. J. 244:* 219 (1987)), and mouse liver (Dailey and Karr,

Biochem. 26: 2697 (1987)). Genes encoding protox have been isolated from two prokaryotic organisms, *Escherichia coli* (Sasarman *et al.*, *Can. J. Microbiol.* 39: 1155 (1993)) and *Bacillus subtilis* (Dailey *et al.*, *J. Biol. Chem.* 269: 813 (1994)). These genes share no sequence similarity; neither do their predicted protein products share any amino acid sequence identity. The *E. coli* protein is approximately 21 kDa, and associates with the cell membrane. The *B. subtilis* protein is 51 kDa, and is a soluble, cytoplasmic activity.

Protox encoding genes have now also been isolated from humans (*see* Nishimura *et al.*, *J. Biol. Chem.* 270(14): 8076-8080 (1995) and plants (International application no. PCT/IB95/00452 filed June 8, 1995, published Dec. 21, 1995 as WO 95/34659).

II. The Protox Gene as a Herbicide Target

The use of herbicides to control undesirable vegetation such as weeds or plants in crops has become almost a universal practice. The relevant market exceeds a billion dollars annually. Despite this extensive use, weed control remains a significant and costly problem for farmers.

Effective use of herbicides requires sound management. For instance, time and method of application and stage of weed plant development are critical to getting good weed control with herbicides. Since various weed species are resistant to herbicides, the production of effective herbicides becomes increasingly important.

Unfortunately, herbicides that exhibit greater potency, broader weed spectrum and more rapid degradation in soil can also have greater crop phytotoxicity. One solution applied to this problem has been to develop crops which are resistant or tolerant to herbicides. Crop hybrids or varieties resistant to the herbicides allow for the use of the herbicides without attendant risk of damage to the crop. Development of resistance can allow application of a herbicide to a crop where its use was previously precluded or limited (*e.g.* to pre-emergence use) due to sensitivity of the crop to the herbicide. For example, U.S. Patent No. 4,761,373 to Anderson *et al.* is directed to plants resistant to various imidazolinone or sulfonamide herbicides. The resistance is conferred by an altered acetohydroxyacid synthase (AHAS) enzyme. U.S.Patent No. 4,975,374 to Goodman *et al.* relates to plant cells and plants containing a gene encoding a mutant glutamine synthetase (GS) resistant to inhibition by herbicides that were known to inhibit GS, *e.g.*

phosphinothricin and methionine sulfoximine. U.S. Patent No. 5,013,659 to Bedbrook *et al.* is directed to plants that express a mutant acetolactate synthase which renders the plants resistant to inhibition by sulfonylurea herbicides. U.S. Patent No. 5,162,602 to Somers *et al.* discloses plants tolerant to inhibition by cyclohexanedione and aryloxyphenoxypropanoic acid herbicides. The tolerance is conferred by an altered acetyl coenzyme A carboxylase(ACCase).

The protox enzyme serves as the target for a variety of herbicidal compounds. The herbicides that inhibit protox include many different structural classes of molecules (Duke *et al.*, *Weed Sci. 39*: 465 (1991); Nandihalli *et al.*, *Pesticide Biochem. Physiol. 43:* 193 (1992); Matringe *et al.*, *FEBS Lett. 245:* 35 (1989); Yanase and Andoh, *Pesticide Biochem. Physiol. 35*: 70 (1989)). These herbicidal compounds include the diphenylethers (e.g. acifluorfen, 5-[2-chloro-4-(trifluoromethyl)phenoxy]-2-nitrobezoic acid; its methyl ester; or oxyfluorfen, 2-chloro-1-(3-ethoxy-4-nitrophenoxy)-4-(trifluorobenzene)), oxidiazoles, (e.g. oxidiazon, 3-[2,4-dichloro-5-(1-methylethoxy)phenyl]-5-(1,1-dimethylethyl)-1,3,4-oxadiazol-2-(3*H*)-one), cyclic imides (e.g. S-23142, *N*-(4-chloro-2-fluoro-5-propargyloxyphenyl)-3,4,5,6-tetrahydrophthalimide; chlorophthalim, *N*-(4-chlorophenyl)-3,4,5,6-tetrahydrophthalimide), phenyl pyrazoles (e.g. TNPP-ethyl, ethyl 2-[1-(2,3,4-trichlorophenyl)-4-nitropyrazolyl-5-oxy]propionate; M&B 39279), pyridine derivatives (e.g. LS 82-556), and phenopylate and its *O*-phenylpyrrolidino- and piperidinocarbamate analogs. Many of these compounds competitively inhibit the normal reaction catalyzed by the enzyme, apparently acting as substrate analogs.

Typically, the inhibitory effect on protox is determined by measuring fluorescence at about 622 to 635 nm, after excitation at about 395 to 410 nM (see, e.g. Jacobs and Jacobs, *Enzyme 28:* 206 (1982); Sherman *et al.*, *Plant Physiol. 97:* 280 (1991)). This assay is based on the fact that protoporphyrin IX is a fluorescent pigment, and protoporphyrinogen IX is nonfluorescent.

The predicted mode of action of protox-inhibiting herbicides involves the accumulation of protoporphyrinogen IX in the chloroplast. This accumulation is thought to lead to leakage of protoporphyrinogen IX into the cytosol where it is oxidized by a peroxidase activity to protoporphyrin IX. When exposed to light, protoporphyrin IX can cause formation of singlet oxygen in the cytosol. This singlet oxygen can in turn lead to the formation of other reactive oxygen species, which can cause lipid peroxidation and membrane disruption leading to rapid cell death (Lee *et al.*, *Plant Physiol. 102*: 881 (1993)).

Not all protox enzymes are sensitive to herbicides which inhibit plant protox enzymes. Both of the protox enzymes encoded by genes isolated from *Escherichia coli* (Sasarman *et al.*, *Can. J. Microbiol. 39:* 155 (1993)) and *Bacillus subtilis* (Dailey *et al.*, *J. Biol. Chem. 269*: 813 (1994)) are resistant to these herbicidal inhibitors. In addition, mutants of the unicellular alga *Chlamydomonas reinhardtii* resistant to the phenylimide herbicide S-23142 have been reported (Kataoka *et al.*, *J. Pesticide Sci. 15:* 449 (1990); Shibata *et al.*, In Research in Photosynthesis. Vol. III, N. Murata, ed. Kluwer:Netherlands. pp. 567-570 (1992)). At least one of these mutants appears to have an altered protox activity that is resistant not only to the herbicidal inhibitor on which the mutant was selected, but also to other classes of protox inhibitors (Oshio *et al.*, Z. *Naturforsch. 48c:* 339 (1993); Sato *et al.*, In ACS Symposium on Porphyric Pesticides. S. Duke, ed. ACS Press: Washington, D.C. (1994)). A mutant tobacco cell line has also been reported that is resistant to the inhibitor S-21432 (Che *et al.*, Z. *Naturforsch. 48c:* 350 (1993).

SUMMARY OF THE INVENTION

The present invention provides isolated DNA molecules and chimeric genes encoding the protoporphyrinogen oxidase (protox) enzyme from soybean and wheat. The sequence of such isolated DNA molecules are set forth in SEQ ID Nos. 11 (soybean) and 9 (wheat).

The present invention also provides modified forms of the plant protoporphyrinogen oxidase (protox) enzyme which are resistant to compounds that inhibit unmodified naturally occurring plant protox enzymes, and DNA molecules coding for such inhibitor-resistant plant protox enzymes. The present invention includes chimeric genes and modified forms of naturally occurring protox genes which can express the inhibitor-resistant plant protox enzymes in plants.

Genes encoding inhibitor-resistant plant protox enzymes can be used to confer resistance to protox-inhibitory herbicides in whole plants and as a selectable marker in plant cell transformation methods. Accordingly, the present invention also includes plants, plant tissues and plant seeds containing plant expressible genes encoding these modified protox enzymes. These plants, plant tissues and plant seeds are resistant to protox-inhibitors at levels which normally are inhibitory to the naturally occurring protox activity in the plant. Plants encompassed by the

invention especially include those which would be potential targets for protox inhibiting herbicides, particularly agronomically important crops such as maize and other cereal crops such as wheat, oats, rye, sorghum, rice, barley, millet, turf and forage grasses, and the like, as well as cotton, sugar cane, sugar beet, oilseed rape, and soybeans.

The present invention is directed further to methods for the production of plants, plant tissues, and plant seeds which produce an inhibitor-resistant form of the plant protox enzyme provided herein. Such plants may be stably transformed with a structural gene encoding the resistant protox, or prepared by direct selection techniques whereby herbicide resistant lines are isolated, characterized and developed.

The present invention is further directed to probes and methods for detecting the presence of genes encoding inhibitor-resistant forms of the plant protox enzyme and quantitating levels of inhibitor-resistant protox transcripts in plant tissue. These methods may be used to identify or screen for plants or plant tissue containing and/or expressing a gene encoding an inhibitor-resistant form of the plant protox enzyme.

DESCRIPTI N OF THE SEQUENCE LISTING

SEQ ID No. 1: DNA coding sequence for an *Arabidopsis thaliana* protox-1 protein.

SEQ ID No. 2: *Arabidopsis thaliana* protox-1 amino acid sequence encoded by SEQ ID No. 1.

SEQ ID No. 3: DNA coding sequence for an *Arabidopsis thaliana* protox-2 protein.

SEQ ID No. 4: *Arabidopsis thaliana* protox-2 amino acid sequence encoded by SEQ ID No.3

SEQ ID No. 5: DNA coding sequence for a maize protox-1 protein.

SEQ ID No. 6: Maize protox-1 amino acid sequence encoded by SEQ ID No. 5

SEQ ID No. 7: DNA coding sequence for a maize protox-2 protein.

SEQ ID No. 8: Maize protox-2 amino acid sequence encoded by SEQ ID No. 7

SEQ ID No. 9: DNA coding sequence for a wheat protox-1 protein.

SEQ ID No. 10: Wheat protox-1 amino acid sequence encoded by SEQ ID No. 9.

SEQ ID No. 11: DNA coding sequence for a soybean protox-1 protein.

SEQ ID No. 12: Soybean protox-1 protein encoded by SEQ ID No. 11.

SEQ ID NO. 13: Promoter sequence from *Arabidopsis thaliana* protox-1 gene.

DETAILED DESCRIPTION OF THE INVENTION

Plant Protox Coding Sequences

In one aspect, the present invention is directed to an isolated DNA molecule which encodes protoporphyrinogen oxidase (referred to herein as "protox"), the enzyme which catalyzes the oxidation of protoporphyrinogen IX to protoporphyrin IX, from soybean and wheat. The DNA coding sequence and corresponding amino acid sequence for a soybean protox enzyme is provided as SEQ ID Nos. 11 and 12, respectively. The DNA coding sequence and correspondir.g amino acid sequence for a wheat protox enzyme is provided as SEQ ID Nos. 9 and 10, respectively.

The DNA coding sequences and corresponding amino acid sequences for protox enzymes from *Arabidopsis thaliana* and maize which have been previously isolated are reproduced herein as SEQ ID Nos. 1-4 (*Arabidopsis*) and SEQ ID Nos 5-8 (maize).

The isolated eukaryotic protox sequences taught by the present invention may be manipulated according to standard genetic engineering techniques to suit any desired purpose. For example, the entire protox sequence or portions thereof may be used as probes capable of specifically hybridizing to protox coding sequences and messenger RNAs. To achieve specific hybridization under a variety of conditions, such probes include sequences that are unique among protox coding sequences and are preferably at least 10 nucleotides in length, and most preferably at least 20 nucleotides in length. Such probes may be used to amplify and analyze protox coding sequences from a chosen organism via the well known process of polymerase chain reaction (PCR). This technique may be useful to isolate additional protox coding sequences from a desired organism or as a diagnostic assay to determine the presence of protox coding sequences in an organism.

Protox specific hybridization probes may also be used to map the location of the native eukaryotic protox gene(s) in the genome of a chosen organism using standard techniques based on the selective hybridization of the probe to genomic protox sequences. These techniques include, but are not limited to, identification of DNA polymorphisms identified or contained within the protox probe sequence, and use of such polymorphisms to follow segregation of the protox gene relative to other markers of known map position in a mapping population derived from self fertilization of a hybrid of two polymorphic parental lines (see e.g. Helentjaris *et al.*, *Plant Mol. Biol. 5:* 109 (1985). Sommer *et al. Biotechniques 12*:82 (1992); D'Ovidio *et al.*, *Plant Mol. Biol. 15:* 169 (1990)). While any eukaryotic protox sequence is contemplated to be useful as a probe for mapping protox genes from any eukaryotic organism, preferred probes are those protox sequences from organisms more closely related to the chosen organism, and most preferred probes are those protox sequences from the chosen organism. Mapping of protox genes in this manner is contemplated to be particularly useful in plants for breeding purposes. For instance, by knowing the genetic map position of a mutant protox gene that confers herbicide resistance, flanking DNA markers can be identified from a reference genetic map (see, *e.g.*, Helentjaris, *Trends Genet. 3:* 217 (1987)). During introgression of the herbicide resistance trait into a new breeding line, these markers can then be used to monitor the extent of protox-linked flanking chromosomal DNA still present in the recurrent parent after each round of back-crossing.

Protox specific hybridization probes may also be used to quantitate levels of protox mRNA in an organism using standard techniques such as Northern blot analysis. This technique may be useful as a diagnostic assay to detect altered levels of protox expression that may be associated with particular adverse conditions such as autosomal dominant disorder in humans characterized by both neuropsychiatric symptoms and skin lesions, have decreased levels of protox activity (Brenner and Bloomer, *New Engl. J. Med. 302:* 765-(1980)).

For recombinant production of the enzyme in a host organism, the protox coding sequence may be inserted into an expression cassette designed for the chosen host and introduced into the host where it is recombinantly produced. The choice of specific regulatory sequences such as promoter, signal sequence, 5' and 3' untranslated sequences, and enhancer, is within the level of skill of the routineer in the art. The resultant molecule, containing the individual elements linked in proper reading frame, may be inserted into a vector capable of being transformed into the host cell. Suitable expression vectors and methods for recombinant production of proteins are well known for host organisms such as *E. coli* (see, *e.g.* Studier and Moffatt, *J. Mol. Biol. 189:* 113 (1986); Brosius, *DNA 8:* 759 (1989)), yeast (see, *e.g.*, Schneider and Guarente, *Meth. Enzymol. 194:* 373 (1991)) and insect cells (see, *e.g.*, Luckow and Summers, *Bio/Technol. 6:* 47 (1988)). Specific examples include plasmids such as pBluescript (Stratagene, La Jolla, CA), pFLAG (International Biotechnologies, Inc., New Haven, CT), pTrcHis (Invitrogen, La Jolla, CA), and baculovirus expression vectors, e.g., those derived from the genome of *Autographica californica nuclear polyhedrosis virus* (AcMNPV). A preferred baculovirus/insect system is pVI11392/Sf21 cells (Invitrogen, La Jolla, CA).

Recombinantly produced eukaryotic protox enzyme is useful for a variety of purposes. For example, it may be used to supply protox enzymatic activity *in vitro*. It may also be used in an *in vitro* assay to screen known herbicidal chemicals whose target has not been identified to determine if they inhibit protox. Such an *in vitro* assay may also be used as a more general screen to identify chemicals which inhibit protox activity and which are therefore herbicide candidates. Recombinantly produced eukaryotic protox enzyme may also be used in an assay to identify inhibitor-resistant protox mutants (*see* International application no. PCT/IB95/00452 filed June 8, 1995, published Dec. 21, 1995 as WO 95/34659, incorporated by reference herein in its entirety). Alternatively, recombinantly produced protox enzyme may be used to further characterize its

association with known inhibitors in order to rationally design new inhibitory herbicides as well as herbicide tolerant forms of the enzyme.

Inhibitor Resistant Plant Protox Enzymes

In another aspect, the present invention teaches simple modifications which can be made to the amino acid sequence of any plant protoporphyrinogen oxidase (referred to herein as "protox") enzyme to yield an inhibitor-resistant form of this enzyme.

The present invention is directed to inhibitor-resistant plant protox enzymes having the modifications taught herein, and to DNA molecules encoding these modified enzymes, and to genes capable of expressing these modified enzymes in plants.

The present invention is further directed to plants, plant tissue and plant seeds tolerant to herbicides that inhibit the naturally occurring protox activity in these plants, wherein the tolerance is conferred by a gene expressing a modified inhibitor-resistant protox enzyme as taught herein. Representative plants include any plants to which these herbicides may be applied for their normally intended purpose. Preferred are agronomically important crops, i.e., angiosperms and gymnosperms such as cotton, soybean, rape, sugar beet, maize, rice, wheat, barley, oats, rye, sorghum, millet, turf, forage, turf grasses and the like.

The modified inhibitor-resistant protox enzymes of the invention have at least one amino acid substitution, addition or deletion relative to their naturally occurring counterpart (i.e. inhibitor-sensitive forms which occur naturally in a plant without being manipulated, either directly *via* recombinant DNA methodology or indirectly *via* selective breeding, etc., by man). Amino acid positions which may be modified to yield an inhibitor-resistant form of the protox enzyme, or enhance inhibitor resistance, are indicated in bold type in Table 1 in the context of plant protox-1 sequences from *Arabidopsis*, maize, soybean and wheat. The skilled artisan will appreciate that equivalent changes may be made to any plant protox gene having a structure sufficiently similar to the protox enzyme sequences shown herein to allow alignment and identification of those amino acids which are modified according to the invention to generate inhibitor-resistant forms of the enzyme. Such additional plant protox genes may be obtained using standard techniques as described in International application no. PCT/IB95/00452 filed June 8,

1995, published Dec. 21, 1995 as WO 95/34659 whose relevant parts are herein incorporated by reference.

DNA molecules encoding the herbicide resistant protox coding sequences taught herein may be genetically engineered for optimal expression in a crop plant. This may include altering the coding sequence of the resistance allele for optimal expression in the crop species of interest. Methods for modifying coding sequences to achieve optimal expression in a particular crop species are well known (see, *e.g.* Perlak *et al.*, *Proc. Natl. Acad. Sci. USA 88:* 3324 (1991); Koziel *et al.*, *Biotechnol. 11:* 194 (1993)).

Genetically engineering a protox coding sequence for optimal expression may also include operably linking the appropriate regulatory sequences (i.e. promoter, signal sequence, transcriptional terminators). Examples of promoters capable of functioning in plants or plant cells (i.e., those capable of driving expression of the associated structural genes such as protox in plant cells) include the cauliflower mosaic virus (CaMV) 19S or 35S promoters and CaMV double promoters; nopaline synthase promoters; pathogenesis-related (PR) protein promoters; small subunit of ribulose bisphosphate carboxylase (ssuRUBISCO) promoters, and the like. Preferred promoters will be those which confer high level constitutive expression or, more preferably, those which confer specific high level expression in the tissues susceptible to damage by the herbicide. Preferred promoters are the rice actin promoter (McElroy *et al.*, *Mol. Gen. Genet. 231*: 150 (1991)), maize ubiquitin promoter (EP 0 342 926; Taylor *et al.*, *Plant Cell Rep. 12:* 491 (1993)), and the Pr-1 promoter from tobacco, *Arabidopsis*, or maize (see U.S. Patent Application Serial No. 08/181,271 to Ryals *et al.*, incorporated by reference herein in its entirety). The promoters themselves may be modified to manipulate promoter strength to increase protox expression, in accordance with art-recognized procedures.

The inventors have also discovered that another preferred promoter for use with the inhibitor-resistant protox coding sequences is the promoter associated with the native protox gene (i.e. the protox promoter; *see* copending, co-owned U.S. Provisional Application entitled "Promoters from Protoporphynnogen Oxidase Genes", filed on the same day as the present application and incorporated by reference herein in its entirety.). The protox promoter sequence from an *Arabidopsis* gene is set forth in SEQ ID No. 13.

Since the protox promoter itself is suitable for expression of inhibitor-resistant protox coding sequences, the modifications taught herein may be made directly on the native protox gene present in the plant cell genome without the need to construct a chimeric gene with heterologous regulatory sequences. Such modifications can be made via directed mutagenesis techniques such as homologous recombination and selected for based on the resulting herbicide-resistance phenotype (*see, e.g.* Example 10, Pazkowski *et al., EMBO J. 7:* 4021-4026 (1988), and U.S. Patent No. 5,487,992, particularly columns 18-19 and Example 8). An added advantage of this approach is that beside containing the native protox promoter, the resulting modified gene will also include any other regulatory elements, such as signal or transit peptide coding sequences, which are part of the native gene.

Signal or transit peptides may be fused to the protox coding sequence in chimeric DNA constructs of the invention to direct transport of the expressed protox enzyme to the desired site of action. Examples of signal peptides include those natively linked to the plant pathogenesis-related proteins, e.g. PR-1, PR-2, and the like. *See, e.g.*, Payne *et al., Plant Mol. Biol. 11*:89-94 (1988). Examples of transit peptides include the chloroplast transit peptides such as those described in Von Heijne *et al., Plant Mol. Biol. Rep. 9:*104-126 (1991); Mazur *et al., Plant Physiol. 85*: 1110 (1987); Vorst *et al., Gene 65*: 59 (1988), and mitochondrial transit peptides such as those described in Boutry *et al., Nature 328:*340-342 (1987). Chloroplast and mitochondrial transit peptides are contemplated to be particularly useful with the present invention as protox enzymatic activity typically occurs within the mitochondria and chloroplast. Most preferred for use are chloroplast transit peptides as inhibition of the protox enzymatic activity in the chloroplasts is contemplated to be the primary basis for the action of protox-inhibiting herbicides (Witkowski and Halling, *Plant Physiol. 87:* 632 (1988); Lehnen *et al., Pestic. Biochem. Physiol. 37*: 239 (1990); Duke *et al., Weed Sci. 39*: 465 (1991)). Also included are sequences that result in localization of the encoded protein to various cellular compartments such as the vacuole. *See, for example*, Neuhaus *et al., Proc. Natl. Acad. Sci. USA 88:* 10362-10366 (1991) and Chrispeels, *Ann. Rev. Plant Physiol. Plant Mol. Biol. 42*: 21-53 (1991). The relevant disclosures of these publications are incorporated herein by reference in their entirety.

Chimeric DNA construct(s) of the invention may contain multiple copies of a promoter or multiple copies of the protox structural genes. In addition, the construct(s) may include coding

sequences for markers and coding sequences for other peptides such as signal or transit peptides, each in proper reading frame with the other functional elements in the DNA molecule. The preparation of such constructs are within the ordinary level of skill in the art.

Useful markers include peptides providing herbicide, antibiotic or drug resistance, such as, for example, resistance to hygromycin, kanamycin, G418, gentamycin, lincomycin, methotrexate, glyphosate, phosphinothricin, or the like. These markers can be used to select cells transformed with the chimeric DNA constructs of the invention from untransformed cells. Other useful markers are peptidic enzymes which can be easily detected by a visible reaction, for example a color reaction, for example luciferase, ß-glucuronidase, or ß-galactosidase.

The recombinant DNA molecules of the invention can be introduced into the plant cell in a number of art-recognized ways. Those skilled in the art will appreciate that the choice of method might depend on the type of plant, i.e. monocot or dicot, targeted for transformation. Suitable methods of transforming plant cells include microinjection (Crossway *et al., BioTechniques 4*:320-334 (1986)), electroporation (Riggt *et al, Proc. Natl. Acad. Sci. USA 83*:5602-5606 (1986), *Agrobacterium* mediated transformation (Hinchee *et al., Biotechnology 6*:915-921 (1988)), direct gene transfer (Paszkowski *et al., EMBO J. 3*:2717-2722 (1984)), ballistic particle acceleration using devices available from Agracetus, Inc., Madison, Wisconsin and Dupont, Inc., Wilmington, Delaware (*see, for example*, Sanford *et al.*, U.S. Patent 4,945,050; and McCabe *et al., Biotechnology 6*:923-926 (1988)), and protoplast transformation/regeneration methods (*see* U.S. Patent No. 5,350,689 issued Sept. 27, 1994 to Ciba-Geigy Corp.). Also see, Weissinger *et al., Annual Rev. Genet. 22*:421-477 (1988); Sanford *et al., Particulate Science and Technology 5*:27-37 (1987)(onion); Christou *et al., Plant Physiol. 87*:671-674 (1988)(soybean); McCabe *et al., Bio/Technology 6*:923-926 (1988)(soybean); Datta *et al., Bio/Technology 8*:736-740 (1990)(rice); Klein *et al., Proc. Natl. Acad. Sci. USA, 85*:4305-4309 (1988)(maize); Klein *et al., Bio/Technology 6*:559-563 (1988)(maize); Klein *et al., Plant Physiol. 91*:440-444 (1988)(maize); Fromm *et al., Bio/Technology 8*:833-839 (1990); and Gordon-Kamm *et al., Plant Cell 2*:603-618 (1990)(maize).

Where a herbicide resistant protox allele is obtained via directed mutation of the native gene in a crop plant or plant cell culture from which a crop plant can be regenerated, it may be moved into commercial varieties using traditional breeding techniques to develop a herbicide

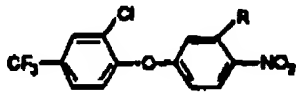
tolerant crop without the need for genetically engineering the modified coding sequence and transforming it into the plant. Alternatively, the herbicide resistant gene may be isolated, genetically engineered for optimal expression and then transformed into the desired variety.

Genes encoding altered protox resistant to a protox inhibitor can also be used as selectable markers in plant cell transformation methods. For example, plants, plant tissue or plant cells transformed with a transgene can also be transformed with a gene encoding an altered protox capable of being expressed by the plant. The thus-transformed cells are transferred to medium containing the protox inhibitor wherein only the transformed cells will survive. Protox inhibitors contemplated to be particularly useful as selective agents are the diphenylethers [e.g. acifluorfen, 5-[2-chloro-4-(trifluoromethyl)phenoxy]-2-nitrobezoic acid; its methyl ester; or oxyfluorfen, 2-chloro-1-(3-ethoxy-4-nitrophenoxy)-4-(trifluorobenzene)], oxidiazoles, (e.g. oxidiazon, 3-[2,4-dichloro-5-(1-methylethoxy)phenyl]-5-(1,1-dimethylethyl)-1,3,4-oxadiazol-2-(3*H*)-one), cyclic imides (e.g. S-23142, *N*-(4-chloro-2-fluoro-5-propargyloxyphenyl)-3,4,5,6-tetrahydrophthalimide; chlorophthalim, *N*-(4-chlorophenyl)-3,4,5,6-tetrahydrophthalimide), phenyl pyrazoles (e.g. TNPP-ethyl, ethyl 2-[1-(2,3,4-trichlorophenyl)-4-nitropyrazolyl-5-oxy]propionate; M&B 39279), pyridine derivatives (e.g. LS 82-556), and phenopylate and its *O*-phenylpyrrolidino- and piperidinocarbamate analogs. The method is applicable to any plant cell capable of being transformed with an altered protox-encoding gene, and can be used with any transgene of interest. Expression of the transgene and the protox gene can be driven by the same promoter functional on plant cells, or by separate promoters.

Modified inhibitor-resistant protox enzymes of the present invention are resistant to herbicides that inhibit the naturally occurring protox activity. The herbicides that inhibit protox include many different structural classes of molecules (Duke *et al.*, *Weed Sci. 39*: 465 (1991); Nandihalli *et al.*, *Pesticide Biochem. Physiol. 43*: 193 (1992); Matringe *et al.*, *FEBS Lett. 245*: 35 (1989); Yanase and Andoh, *Pesticide Biochem. Physiol. 35*: 70 (1989)), including the diphenylethers [e.g. acifluorifen, 5-[2-chloro-4-(trifluoromethyl)phenoxy]-2-nitrobezoic acid; its methyl ester; or oxyfluorfen, 2-chloro-1-(3-ethoxy-4-nitrophenoxy)-4-(trifluorobenzene)], oxidiazoles (e.g. oxidiazon, 3-[2,4-dichloro-5-(1-methylethoxy)phenyl]-5-(1,1-dimethylethyl)-1,3,4-oxadiazol-2-(3*H*)-one), cyclic imides (e.g. S-23142, *N*-(4-chloro-2-fluoro-5-propargyloxyphenyl)-3,4,5,6-tetrahydrophthalimide; chlorophthalim, *N*-(4-chlorophenyl)-3,4,5,6-

tetrahydrophthalimide), phenyl pyrazoles (e.g. TNPP-ethyl, ethyl 2-[1-(2,3,4-trichlorophenyl)-4-nitropyrazolyl-5-oxy]propionate; M&B 39279), pyridine derivatives (e.g. LS 82-556), and phenopylate and its *O*-phenylpyrrolidino- and piperidinocarbamate analogs.

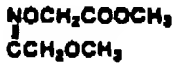
The diphenylethers of particular significance are those having the general formula



(Formula I)

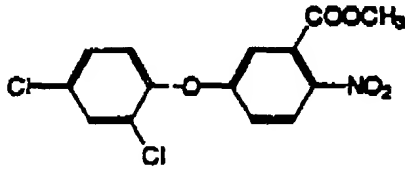
wherein R equals -COONa (Formula II), -$CONHSO_2CH_3$ (Formula III) or -$COOCH_2COOC_2H_5$ (Formula IV; *see* Maigrot *et al., Brighton Crop Protection Conference-Weeds:* 47-51 (1989)).

Additional diphenylethers of interest are those where R equals:



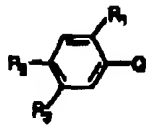
(Formula IVa; *see* Hayashi *et al., Brighton Crop Protection Conference-Weeds:* 53-58 (1989)).

An additional diphenylether of interest is one having the formula:



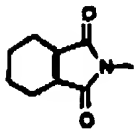
(Formula IVb; bifenox, *see* Dest *et al., Proc. Northeast Weed Sci. Conf.* 27: 31 (1973)).

Also of significance are the class of herbicides known as imides, having the general formula

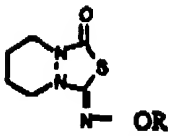


(Formula V)

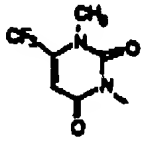
wherein Q equals



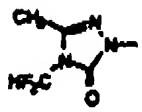
OR



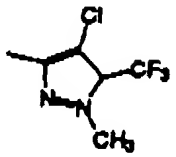
OR



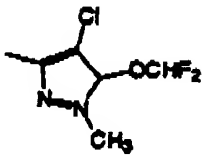
OR



(Formula VI) (Formula VII) (Formula VIII) (Formula IX)



OR

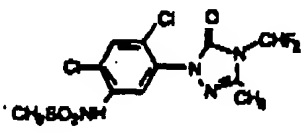


OR

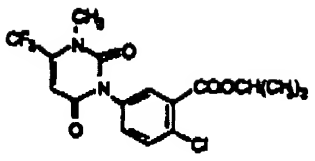
(Formula IXa) (Formula IXb)

(*see* Hemper *et al.* (1995) *in* "Proceedings of the Eighth International Congress of Pesticide Chemistry", Ragdale *et al.*, eds., Amer. Chem. Soc, Washington, D.C., pp.42-48 (1994));

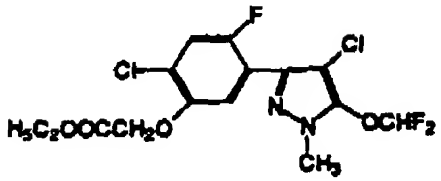
and R_1 equals H, Cl or F, R_2 equals Cl and R_3 is an optically substituted ether, thioether, ester, amino or alkyl group. Alternatively, R_2 and R_3 together may form a 5 or 6 membered heterocyclic ring. Examples of imide herbicides of particular interest are



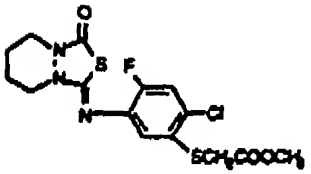
(Formula X)



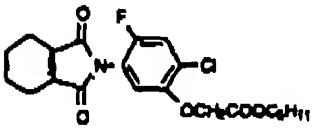
(Formula XI)



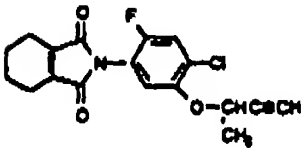
(Formula XII)
(*see* Miura *et al.*, *Brighton Crop Protection Conference-Weeds:* 35-40 (1993))



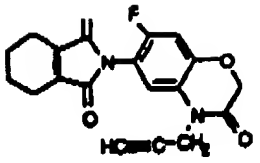
(Formula XIII)



(Formula XIV)



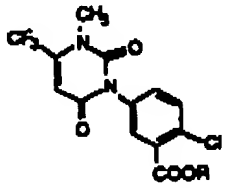
(Formula XV)



(Formula XVI)

The herbicidal activity of the above compounds is described in the *Proceedings of the 1991 Brighton Crop Protection Conference, Weeds* (British Crop Protection Council) (Formulae X and XVI), *Proceedings of the 1993 Brighton Crop Protection Conference, Weeds* (British Crop Protection Council) (Formulae XII and XIII), U.S. Patent No. 4,746,352 (Formula XI) and *Abstracts of the Weed Science Society of America* vol. 33, pg. 9 (1993)(Formula XIV).

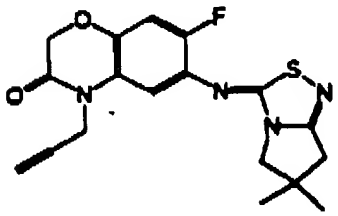
The most preferred imide herbicides are those classified as aryluracils and having the general formula



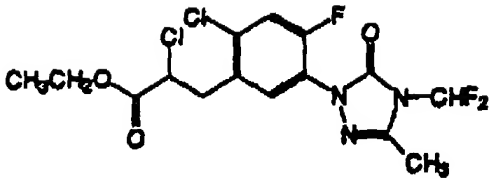
(Formula XVII)

wherein R signifies the group (C_{2-4}-alkenyloxy)carbonyl-C_{1-4}-alkyl, as disclosed in U.S. Patent No. 5,183,492, herein incorporated by reference.

Also of significance are herbicides having the general formula:

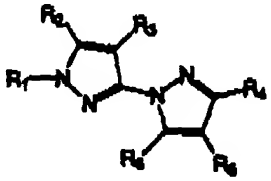


(Formula XVIII; thiadiazimin)
(*see* Weiler *et al.*, *Brighton Crop Protection Conference-Weeds*, pp. 29-34 (1993));



(Formula XIX: carfentrazone)
(*see* Van Saun *et al.*, *Brighton Crop Protection Conference-Weeds:* pp. 19-22 (1993));

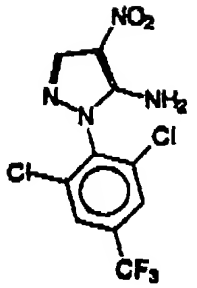
N-substituted pyrazoles of the general formula:



(Formula XX)

wherein R_1 is C_1-C_6-alkyl, optionally substituted by one or more halogen atoms;
R_2 is hydrogen, or a C_1-C_6-alkoxy, each of which is optionally substituted by one or more halogen atoms, or
R_1 and R_2 together from the group -$(CH_2)_n$-X-, where X is bound at R_2;
R_3 is hydrogen or halogen,
R_4 is hydrogen or C_1-C_4-alkyl,
R_5 is hydrogen, nitro, cyano or the group -$COOR_6$ or -$CONR_7R_8$, and
R_6 is hydrogen, C_1-C_6-alkyl, C_2-C_6-alkenyl or C_2-C_6-alkynyl;
(*see* international patent publications WO 94/08999, WO 93/10100, and U. S. Patent No. 5,405,829 assigned to Schering);

N-phenylpyrazoles, such as:



(Formula XXI; nipyraclofen)
(*see* page 621 of "The Pesticide Manual", 9th ed., ed. by C.R. Worthing, British Crop Protection Council, Surrey (1991));

and 3-substituted-2-aryl-4,5,6,7-tetrahydroindazoles (Lyga *et al. Pesticide Sci.* 42:29-36 (1994)).

Levels of herbicide which normally are inhibitory to the activity of protox include application rates known in the art, and which depend partly on external factors such as environment, time and method of application. For example, in the case of the imide herbicides represented by Formulae V through IX, and more particularly those represented by Formulae X through XVII, the application rates range from 0.0001 to 10 kg/ha, preferably from 0.005 to 2 kg/ha. This dosage rate or concentration of herbicide may be different, depending on the desired action and particular compound used, and can be determined by methods known in the art.

The invention will be further described by reference to the following detailed examples. These examples are provided for purposes of illustration only, and are not intended to be limiting unless otherwise specified.

EXAMPLES

Standard recombinant DNA and molecular cloning techniques used here are well known in the art and are described by T. Maniatis, E. F. Fritsch and J. Sambrook, Molecular Cloning: A Laboratory manual, Cold Spring Harbor laboratory, Cold Spring Harbor, NY (1982) and by T.J.

Silhavy, M.L. Berman, and L.W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984).

EXAMPLE 1: Isolation of a wheat Protox-1 cDNA based on sequence homology to a maize Protox-1 coding sequence

Total RNA prepared from *Triticum aestivum* (cv Kanzler) was submitted to Clontech for custom cDNA library construction in the Lambda Uni-Zap vector. Approximately 50,000 pfu of the cDNA library was plated at a density of approximately 5,000 pfu per 10 cm Petri dish and duplicate filter lifts were made onto nitrocellulose membranes (Schleicher and Schuell). The plaque lifts were probed with the maize Protox-1 cDNA (SEQ ID No 5; *see* Example 2 of International application no. PCT/IB95/00452 filed June 8, 1995, published Dec. 21, 1995 as WO 95/34659) labeled with 32P-dCTP by the random priming method (Life Technologies). Hybridization and wash conditions were at 50° C as described in Church and Gilbert, *Proc. Natl. Acad. Sci. USA 81:* 1991-1995 (1984). Positively hybridizing plaques were purified and in vivo excised into pBluescript plasmids. The sequences of the cDNA inserts were determined by the chain termination method using dideoxy terminators labeled with fluorescent dyes (Applied Biosystems, Inc.). The longest wheat cDNA obtained from initial screening efforts, designated "wheat Protox-1", was 1489 bp in length. Wheat Protox-1 lacks coding sequence for the transit peptide plus approximately 126 amino acids of the mature coding sequence based on comparison with the other known plant protox peptide sequences (Table 1). This partial wheat protein sequence is 90% identical (94% similar) to the maize Protox-1 protein.

A second screen was performed to obtain a longer wheat protox cDNA. For this screen a *Triticum aestivum* (cv Kanzler) cDNA library was prepared internally using the lambda Uni-Zap vector. Approximately 200,000 pfu of the cDNA library was screened as indicated above, except that the wheat protox-1 cDNA was used as a probe and hybridization and wash conditions were at 65° C instead of 50° C. The longest wheat cDNA obtained from this screening effort, designated "wheat Protox-1a", was 1811bp in length. The nucleotide sequence of this cDNA and the amino acid sequence it encodes is set forth in SEQ. ID. Nos. 9 and 10, respectively. Based on comparison with the other known plant protox peptide sequences and with corresponding genomic sequence, this cDNA is either full-length or missing only a few transit peptide codons.

This wheat protein sequence is 91% identical (95% similar) to the maize Protox-1 protein sequence set forth in SEQ ID No. 6.

EXAMPLE 2: Isolation of a soybean Protox-1 cDNA based on sequence homology to an Arabidopsis Protox-1 coding sequence

A Lambda Uni-Zap cDNA library prepared from soybean (v Williams 82, epicotyls) was purchased from Stratagene. Approximately 50,000 pfu of the library was plated at a density of approximately 5,000 pfu per 10 cm Petri dish and duplicate filter lifts were made onto Colony/Plaque Screen membranes (NEN Dupont). The plaque lifts were probed with the Arabidopsis Protox-1 cDNA (SEQ ID No. 1; *see* Example 1 of International application no. PCT/IB95/00452 filed June 8, 1995, published Dec. 21, 1995 as WO 95/34659) labeled with 32P-dCTP by the random priming method (Life Technologies). Hybridization conditions were 7% sodium dodecyl sulfate (SDS), 0.5 M NaPO4 pH 7.0, 1 mM EDTA at 50 C. Wash conditions were 2X SSC, 1% SDS at 50 C. Positively hybridizing plaques were purified and in vivo excised into pBluescript plasmids. The sequence of the cDNA inserts was determined by the chain termination method using dideoxy terminators labeled with fluorescent dyes (Applied Biosystems, Inc.). The longest soybean cDNA obtained, designated "soybean Protox-1", is full-length based on comparison with the other known plant protox peptide sequences (Table 1). Soybean Protox-1 is 1847 bp in length and encodes a protein of 58.8 kD. The N-terminal peptide sequence has features characteristic of a chloroplast transit peptide of approximately 65 amino acids. The soybean protein is 78% identical (87% similar) to the Arabidopsis Protox-1 protein.

Soybean Protox-1, in the pBluescript SK vector, was deposited December 14, 1995 as pWDC-12 (NRRL #B-21516).

An alignment of the predicted amino acid sequences of the respective proteins encoded by the sequences shown in SEQ ID NOS: 2 and 6 are set forth in Table 1. An alignment of the predicted amino acid sequences of the respective proteins encoded by the sequences shown in SEQ ID NOS: 4 and 8 are set forth in Table 2.

TABLE 1

Comparison of Protox-1 Amino Acid Sequences from Arabidopsis ("Protox-1"; SEQ ID No. 2), Maize ("Maizeprotox-1"; SEQ ID No. 6), Wheat ("Wheatpt1"; SEQ ID NO. 10) and Soybean ("Soybeanpt"; SEQ ID NO. 12)

Identical residues are denoted by the vertical bar between the two sequences. Alignment is performed using the GAP program described in Deveraux *et al., Nucleic Acids Res.* 12:387-395 (1984). Positions which may be modified according to the teachings herein to confer or enhance inhibitor resistance are shown in bold type.

```
Mzprotox-1  .......... .......... .......... .......... ..........
  Wheatpt1  .......... .......... .......... .......... ..........
 Soybeanpt  MVSVPNEILP PPNQTLLRPS LHSPTSFFTS PTRKFPRSRP WPILRCSIAE
   Protox-1 .......... .MELSLLRPT TQSLLPSPSK PKLALNVTKP LR.LRCSVAG

                                                              100
Mzprotox-1  .......... ..MS....AD CVVVGGGISG LCTAQALATR HGVG..DVLV
  Wheatpt1  .......... .......... .......... .......... ..........
 Soybeanpt  ESTASPPKTR DSAP....VD CVVVGGGVSG LCIAQALATK HANA..NVVV
   Protox-1 GPTVGSSKIE GGGGTTITTD CVIVGGGISG LCIAQALATK HPDAAPNLIV

                                                              150
Mzprotox-1  TEARARPGGN ITTVEKPEEG YLWEEGPNSF QPSDPVLTMA VDSGLKDDLV
  Wheatpt1  .......... .......... .......... .......... ..........
 Soybeanpt  TEARDRVGGN ITTHER..DG YLWEEGPNSF QPSDPKLTMV VDSGLKDELV
   Protox-1 TEARDRVGGN IIT..REEMG FLWEEGPNSF QPSDPMLTMV VDSGLKDDLV

                                                              200
Mzprotox-1  PGDPNAPRFV LWEGKLRPVP SKPADLPFFD LMSIPGKLRA GLGALGIRPP
  Wheatpt1  .......... .......... .......... .......... .......RPP
 Soybeanpt  LGDPDAPRFV LWNRKLRPVP GKLTDLPFFD LMSIQGKIRA GFGALGIRPP
   Protox-1 LGDPTAPRFV LWEGKLRPVP SKLTDLPFFD LMSIQGKIRA GFGALGIRPP

                                                              250
Mzprotox-1  PPGREESVEE FVRRNLGAEV FERLIEPFCS GVYAGDPSKL SMKAAFGKVW
  Wheatpt1  PPGREESVEE FVRRNLGAEV FERLIEPFCS GVYAGDPSKL SMKAAFGKVW
 SoyLeanpt  PPGHEESVEE FVRRNLGDEV FERLIEPFCS GVYAGDPSKL SMKAAFGKVW
   Protox-1 PPGREESVEE FVRRNLGDEV FERLIEPFCS GVYAGDPSKL SMKAAFGKVW

                                                              300
Mzprotox-1  RLEETGGSII GGTIKTIQER SKNPKPPRDA RLPKPKGQTV ASFRKGLAML
  Wheatpt1  RLEEIGGSII GGTIKAIQDK GKNPKPPRDP RLPAPKGQTV ASFRKGLAML
 Soybeanpt  KLEKNGGSII GGTFKAIQER NGASKPPRDP RLPKPKGQTV GSFRKGLTML
   Protox-1 KLEQNGGSII GGTFKAIQER KNAPKAERDP RLPKPQGQTV GSFRKGLRML

                                                              350
Mzprotox-1  PNAITSSLGS KVKLSWKLTS ITKSDNKGYV LEYETPEGVV SVQAKSVIMT
  Wheatpt1  PNAIASRLGS KVKLSWKLTS ITKADNQGYV LGYETPEGLV SVQAKSVIMT
 Soybeanpt  PDAISARLGN KVKLSWKLSS ISKLDSGEYS LTYETPEGVV SLQCKTVVLT
   Protox-1 PEAISARLGS KVKLSWKLSG ITKLESGGYN LTYETPDGLV SVQSKSVVMT
```

TABLE 1

(Continued)

```
                351                                                 400
Mzprotox-1      IPSYVASNIL RPLSSDAADA LSRFYYPPVA AVTVSYPKEA IRKECLIDGE
Wheatptl        TPSYVASDIL RPLSIDAADA LSKFYYPPVA AVTVSYPKEA IRKECLIDGE
Soybeanpt       IPSYVASTLL RPLSAAAADA LSKFYYPPVA AVSISYPKEA IRSECLIDGE
Protox-1        VPSHVASGLL RPLSESAANA LSKLYYPPVA AVSISYPKEA IRTECLIDGE

                401                                                 450
Mzprotox-1      LQGFGQLHPR SQGVETLGTI YSSSLFPNRA PDGRVLLLNY IGGATNTGIV
Wheatptl        LQGFGQLHPR SQGVETLGTI YSSSLFPNRA PAGKVLLLNY IGGSTNTGIV
Soybeanpt       LKGFGQLHPR SQGVETLGTI YSSSLFPNRA PPGRVLLLNY IGGATNTGIL
Protox-1        LKGFGQLHPR TQGVETLGTI YSSSLFPNRA PPGRILLLNY IGGSTNTGIL

                451                                                 500
Mzprotox-1      SKTESELVEA VDRDLRKMLI NSTAVDPLVL GVRVWPQAIP QFLVGHLDLL
Wheatptl        SKTESDLVGA VDRDLRKMLI NPRAADPLAL GVRVWPQAIP QFLIGHLDRL
Soybeanpt       SKTDSKLVET VDRDLRKILI NPNAQDPFVV GVRLWPQAIP QFLVGHLDLL
Protox-1        SKSEGELVEA VDRDLRKMLI KPNSTDPLKL GVKVWPQAIP QFLVGHFDIL

                501                                                 550
Mzprotox-1      EAAKAALDRG GYDGLFLGGN YVAGVALGRC VEGAYESASQ ISDFLTKYAY
Wheatptl        AAAKSALGQG GYDGLFLGGN YVAGVALGRC IEGAYESASQ VSDFLTKYAY
Soybeanpt       DVAKASIRNT GFEGLFLGGN YVSGVALGRC VEGAYEVAAE VNDFLTNRVY
Protox-1        DTAKSSLTSS GYEGLFLGGN YVAGVALGRC VEGAYETAIE VNNFMSRYAY

                551
Mzprotox-1      K*
Wheatptl        K*
Soybeanpt       K*
Protox-1        K*
```

TABLE 2

Comparison of the Arabidopsis (SEQ ID No. 4) and Maize (SEQ ID NO. 8) Protox-2 Amino Acid Sequences

Percent Similarity: 75.889 Percent Identity: 57.905
Protox-2.Pep x Mzprotox-2.Pep

```
          1 ..........................MASGAVAD.HQIEAVSGKRVAV  21
                                      .|  |:|: .:  |..::.|||
          1 MLALTASASSASSHPYRHASAHTRAPRLRAVLAMAGSDDPRAAPARSVAV  50

         22 VGAGVSGLAAAYKLKSRGLNVTVFEADGRVGGKLRSVMQNGLIWDEGANT  71
            ||||||||||||:|: .|:|||||||.:|.|||:|.  :.|::|||||||
         51 VGAGVSGLAAAYRLRQSGVNVTVFEAADRAGGKIRTNSEGGFVWDEGANT 100

         72 MTEAEPEVGSLLDDLGLREKQQFPISQKKRYIVRNGVPVMLPTNPIELVT 121
            |||:| |.:.|:|||||.:|||:| ||.||||::|.|.::|.:||.|:.
        101 MTEGEMEASRLIDDLGLQDKQQYPNSQHKRYIVKDGAPALIPSDPISLMK 150

        122 SSVLSTQSKFQILLEPFLWKK....KSSKVSDASAEESVSEFFQRHFGQE 167
            ||||||.||:.:::||||:|| .|:|||:. .|||:.| :|||||.|
        151 SSVLSTKSKIALFFEPFLYKKAMTRNSGKVSEEHLSESVGSFCERHFGRE 200

        168 VVDYLIDPFVGGTSAADPDSLSMKHSFPDLWNVEKSFGSIIVGAIRTKPA 217
            ||||::|||||||||:|||::|.||.|||:|:.:||:|||| .|:|
        201 VVDYFVDPFVAGTSAGDPESLSIRHAFPALWNLERKYGSVIVGAILSKLA 250

        218 AKGGKSRDTKSSPGTKKGSRGSFSFKGGMQILPDTLCKSLSNDEINLDSK 267
            |||:. :. ..|.|.:::.|.||||.|||| | :.| ..::.|::.|:..
        251 AKGDPVKTRHDSSGKRNNRRVSFSFHGGMQSLINALHNEVGDDNVKLGTE 300

        268 VLSLS..YNSGSRQLNWSLSCVSHNETQRQ...NFHYDAVIKTAPLCNVK 312
            ||||. :::.. :.||:|. |.:..:::   |. :|||||||||:||:
        301 VLSLACTFDGVPALGRWSISVDSKDSGDKDLASNQTFDAVIKTAPLSNVR 350

        313 EMKVMKGGQPFQLNFLPEINYMPLSVLITTFTKEKVKRPLDGFGVLIPSK 362
             ||. |||.|. |:|||.::|:|||::|.|.:|.||:|||||||||| |
        351 RMKFTKGGAPVVLDFLPKNDYLPLSLMVTAFKKDDVKKPLEGFGVLIPYK 400

        363 E.QKHGFKTLGTLFSSMMFPDRSPSDVHLYTTFIGGSRNQELAKASTDEL 411
            | ||||:|||||||||||||||.|.| .|||||:|||:|.:|| |.|. |
        401 EQQKHGLKTLGTLFSSMMFPDRAPDDQYLYTTFVGGSHNRDLAGAPTSIL 450

        412 KQVVTSDLQRLLGVEGEPVSVNHYYWRKAFPLYDSSYDSVMEAIDKMEND 461
            ||:|||||.:|||||||:|. |.| || .|||||: .|.||:|||:|||.:
        451 KQLVTSDLKKLLGVEGQPTFVKHVYWGNAFPLYGHDYSSVLEAIEKMEKN 500

        462 LPGFFYAGNHRGGLSVGKSIASGCKAADLVISYLESCSMDKKPNDSL* 509
            ||||||||| ::||.||. ||||:|||||.|||||| ......:
        501 LPGFFYAGNSKDGLAVGSVIASGSKAADLAISYLESHTKNNNSH*... 545
```

EXAMPLE 3: Demonstration of plant protox clone sensitivity to protox inhibitory herbicides in a bacterial system.

Liquid cultures of Protox-1/SASX38, Protox-2/SASX38 and pBluescript/XL1-Blue were grown in L amp^{100}. One hundred microliter aliquots of each culture were plated on L amp^{100} media containing various concentrations (1.0nM-10mM) of a protox inhibitory aryluracil herbicide of formula XVII. Duplicate sets of plates were incubated for 18 hours at 37° C in either low light or complete darkness.

The protox$^+$ *E. coli* strain . 1-Blue showed no sensitivity to the herbicide at any concentration, consistent with repor d resistance of the native bacterial enzyme to similar herbicides. The Protox-1/SASX38 was clearly sensitive, with the lawn of bacteria almost entirely eliminated by inhibitor concentrations as low as 10nM. The Protox-2/SASX38 was also sensitive, but only at a higher concentration (10μM) of the herbicide. The effect of the herbicide on both plant protox strains was most dramatic in low light, but was also apparent on plates maintained entirely in the dark. The toxicity of the herbicide was entirely eliminated by the addition of 20mg/ml hematin to the plates.

The different herbicide tolerance between the two plant Protox strains is likely the result of differential expression from these two plasmids, rather than any inherent difference in enzyme sensitivity. Protox-1/SASX38 grows much more slowly than Protox-2/SASX38 in any heme-deficient media. In addition, the MzProtox-2/SASX38 strain, with a growth rate comparable to Arab Protox-1/SASX38, is also very sensitive to herbicide at the lower (10-100nM) concentrations. Initial characterization of the yeast Protox-3 clone indicated that it also is herbicide sensitive.

EXAMPLE 4: Selecting for plant protox genes resistant to protox-inhibitory herbicides in the E. coli expression system

An Arabidopsis thaliana (Landsberg) cDNA library in the plasmid vector pFL61 (Minet *et al., Plant J. 2*:417-422 (1992) was obtained and amplified. The E. coli hemG mutant SASX38 (Sasarman *et al., J. Gen. Microbiol. 113*:297(1979)) was obtained and maintained on L media containing 20ug/ml hematin (United States Biochemicals). The plasmid library was transformed into SASX38 by electroporation using the Bio-Rad Gene Pulser and the manufacturer's conditions. The electroporated cells were plated on L agar containing 100ug/ml ampicillin at a density of approximately 500,000 transformants/10cm plate. The cells were then incubated at 37^0

C for 40 hours in low light and selected for the ability to grow without the addition of exogenous heme. Heme prototrophs were recovered at a frequency of $400/10^7$ from the pFL61 library. Sequence analysis of twenty-two complementing clones showed that nine are of the type designated "Protox-1," the protox gene expected to express a chloroplastic protox enzyme.

The pFL61 library is a yeast expression library, with the Arabidopsis cDNAs inserted bidirectionally. These cDNAs can also be expressed in bacteria. The protox cDNAs apparently initiate at an in-frame ATG in the yeast PGK 3' sequence approximately 10 amino acids 5' to the NotI cloning site in the vector and are expressed under control of the lacZ promoter 300bp further upstream. Because Protox-1 cDNAs that included significant portions of a chloroplast transit sequence inhibited the growth of the E. coli SASX38 strain, the clone with the least amount of chloroplast transit sequence attached was chosen for mutagenesis/herbicide selection experiments. This clone, pSLV19, contains only 17 amino acids of the putative chloroplast transit peptide, with the DNA sequence beginning at bp 151 of the Arabidopsis Protox-1 cDNA (SEQ ID NO. 1).

The plasmid pSLV19 was transformed into the random mutagenesis strain XL1-Red (Stratagene, La Jolla, CA). The transformation was plated on L media containing 50ug/ml ampicillin and incubated for 48 hours at 37°C. Lawns of transformed cells were scraped from the plates and plasmid DNA prepared using the Wizard Megaprep kit (Promega, Madison, WI). Plasmid DNA isolated from this mutator strain is predicted to contain approximately one random base change per 2000 nucleotides (see Greener *et al., Strategies 7(2)*:32-34 (1994).

The mutated plasmid DNA was transformed into the hemG mutant SASX38 (Sasarman *et al., J. Gen. Microbiol. 113*:297 (1979) and plated on L media containing various concentrations of protox-inhibiting herbicide. The plates were incubated for 2 days at 37°C. Plasmid DNA was isolated from all colonies that grew in the presence of herbicide concentrations that effectively killed the wild type strain. The isolated DNA was then transformed into SASX38 and plated again on herbicide to ensure that the resistance observed was plasmid-borne. The protox coding sequence from plasmids passing this screen was excised by NotI digestion, recloned into an unmutagenized vector, and tested again for the ability to confer herbicide tolerance. The DNA sequence of protox cDNAs that conferred herbicide resistance was then determined and mutations identified by comparison with the wild type Arabidopsis Protox-1 sequence (SEQ ID NO. 1).

A single coding sequence mutant was recovered from the first mutagenesis experiment. This mutant leads to enhanced herbicide "resistance" only by increasing growth rate. It contains a C to A mutation at nucleotide 197 at SEQ ID NO. 1 in the truncated chloroplast transit sequence of pSLV19, converting an ACG codon for Threonine to an AAG codon for Lysine at amino acid 56 of SEQ ID NO. 2, and resulting in better complementation of the bacterial mutant. This plasmid also contains a silent coding sequence mutation at nucleotide 1059, with AGT (Ser) changing to AGC (Ser). This plasmid was designated pMut-1.

The pMut-1 plasmid was then transformed into the mutator XL1-Red strain as described above and the mutated DNA was isolated and plated on an herbicide concentration that is lethal to the unmutagenized pMut-1 protox gene. Herbicide tolerant colonies were isolated after two days at 37 C and analyzed as described above. Multiple plasmids were shown to contain herbicide resistant protox coding sequences. Sequence analysis indicated that the resistant genes fell into three classes. One resistance mutation identified was a C to T change at nucleotide 689 in the Arabidopsis Protox-1 sequence set forth in SEQ ID NO. 1. This change converts a GCT codon for alanine at amino acid 220 of SEQ ID NO. 2 to a GTT codon for valine, and was designated pAraC-1Val.

A second class of herbicide resistant mutant contains an A to G change at nucleotide 1307 in the Arabidopsis Protox-1 sequence. This change converts a TAC codon for tyrosine at amino acid 426 to a TGC codon for cysteine, and was designated pAraC-2Cys.

A third resistant mutant has a G to A change at nucleotide 691 in the Arabidopsis Protox-1 sequence. This mutation converts a GGT codon for glycine at amino acid 221 to an AGT codon for serine at the codon position adjacent to the mutation in pAraC-1. This plasmid was designated pAraC-3Ser.

Resistant mutant pAraC-2Cys, in the pMut-1 plasmid, was deposited on November 14, 1994 under the designation pWDC-7 with the Agricultural Research Culture Collection and given the deposit designation NRRL #21339N.

EXAMPLE 5: Additional herbicide-resistant codon substitutions at positions identified in the random screen

The amino acids identified as herbicide resistance sites in the random screen are replaced by other amino acids and tested for function and for herbicide tolerance in the bacterial system. Oligonucleotide-directed mutagenesis of the Arabidopsis Protox-1 sequence is performed using the Transformer Site-Directed Mutagenesis Kit (Clontech, Palo Alto, CA). After amino acid changes are confirmed by sequence analysis, the mutated plasmids are transformed into SASX38 and plated on L-amp100 media to test for function and on various concentrations of protox-inhibiting herbicide to test for tolerance.

This procedure is applied to the alanine codon at nucleotides 688-690 (amino acid 220 of SEQ ID No. 2) and to the tyrosine codon at nucleotides 1306-1308 (amino acid 426 of SEQ ID No. 2) of the Arabidopsis Protox-1 sequence (SEQ ID NO. 1). The results demonstrate that the alanine codon at nucleotides 688-690 can be changed to a codon for valine, threonine, leucine, cysteine or isoleucine to yield an herbicide-resistant protox enzyme which retains function. The results further demonstrate that the tyrosine codon at nucleotides 1306-1308 can be changed to a codon for cysteine, isoleucine, leucine, threonine or methionine to yield an herbicide-resistant protox enzyme which retains function.

EXAMPLE 6: Isolation of additional mutations that increase enzyme function and/or herbicide tolerance of previously identified resistant mutants

Plasmids containing herbicide resistant protox genes are transformed into the mutator strain XL1-Red and mutated DNA is isolated as described above. The mutated plasmids are transformed into SASX38 and the transformants are screened on herbicide concentrations sufficient to inhibit growth of the original "resistant" mutant. Tolerant colonies are isolated and the higher tolerance phenotype is verified as being coding sequence dependent as described above. The sequence of these mutants is determined and mutations identified by comparison to the progenitor sequence.

This procedure was applied to the pAraC-1Val mutant described above. The results demonstrate that the serine codon at amino acid number 305 (SEQ ID NO. 2) can be changed to a

codon for leucine to yield an enzyme with higher tolerance to protox-inhibiting herbicides than the pAraC-1Val mutant alone. This second site mutation is designated AraC305Leu. The same results are demonstrated for the threonine codon at amino acid 249, where a change to either isoleucine or to alanine leads to a more tolerant enzyme . These changes are designated AraC249Ile and AraC249Ala, respectively.

The procedure was also applied to the pAraC-2Cys mutant described above. The results demonstrate that the proline codon at amino acid 118 (SEQ ID NO. 2) can be changed to a codon for leucine to yield an enzyme with higher tolerance to protox-inhibiting herbicides than the pAraC-2Cys mutant alone. This mutation is designated AraC118Leu. The same results are demonstrated for the serine codon at amino acid 305, where a change to leucine leads to a more tolerant pAraC-2Cys enzyme. This change was also isolated with the pAraC-1Val mutant as described above and is designated AraC305Leu. Additional mutations that enhance the herbicide resistance of the pAraC-2Cys mutant include an asparagine to serine change at amino acid 425, designated AraC425Ser, and a tyrosine to cysteine at amino acid 498, designated AraC498Cys.

These changes are referred to as "second site" mutations because they were found not sufficient to confer herbicide tolerance alone, but rather enhance the function and/or the herbicide tolerance of an already mutant enzyme. This does not preclude the possibility that other amino acid substitutions at these sites could suffice to produce a herbicide tolerant enzyme since exhaustive testing of all possible replacements has not been performed.

EXAMPLE 7: Identification of additional sites in the maize Protox-1 gene that can be mutated to give herbicide tolerance

The pMut-1 Arabidopsis Protox -1 plasmid described above is very effective when used in mutagenesis/screening experiments in that it gives a high frequency of genuine coding sequence mutants, as opposed to the frequent up-promoter mutants that are isolated when other plasmids are used. In an effort to create an efficient plasmid screening system for the maize Protox-1 cDNA, the maize cDNA was engineered into the pMut-1 vector in approximately the same sequence context as the Arabidopsis cDNA. Using standard methods of overlapping PCR fusion, the 5' end of the pMut-1 Arabidopsis clone (including 17 amino acids of chloroplast transit peptide with one missense mutation as described above) was fused to the maize Protox-1 cDNA

sequence starting at amino acid number 16 (SEQ ID NO. 6) of the maize sequence. The 3' end of the maize cDNA was unchanged. NotI restriction sites were placed on both ends of this fusion, and the chimeric gene was cloned into the pFL61 plasmid backbone from pMut-1. Sequence analysis revealed a single nucleotide PCR-derived silent mutation which converts the ACG codon at nucleotides 752-754 (SEQ ID NO. 5) to an ACT codon, both of which encode threonine. This chimeric Arab-maize Protox-1 plasmid is designated pMut-3.

The pMut-3 plasmid was transformed into the mutator XL1-Red strain as described above and the mutated DNA was isolated and plated on a herbicide concentration that was lethal to the unmutagenized pMut-3 maize protox gene. Herbicide tolerant colonies were isolated after two days at 37°C and analyzed as described above. This analysis revealed multiple plasmids containing herbicide resistant protox coding sequences. Sequence analysis showed 5 single base changes that individually result in a herbicide tolerant maize Protox-1 enzyme. Three of these mutations correspond to amino acid changes previously shown to confer tolerance at the corresponding position in the *Arabidopsis* Protox-1 gene. Two of the three are pMzC-1Val and pMzC-1Thr, converting the alanine (GCT) at amino acid 166 (SEQ ID NO. 6) to either valine (GAT) or to threonine (ACT). This position corresponds to the pAraC-1 mutations described above. The third analogous change converts the glycine (GGT) at amino acid 167 to Serine (AGT), corresponding to the AraC-3Ser mutation described above. These results serve to validate the expectation that herbicide-tolerant mutations identified in one plant protox gene will also confer herbicide tolerance in an equivalent plant protox gene from another species.

Two of the mutations isolated from the maize Protox-1 screen result in amino acid changes at residues not previously identified as herbicide resistance sites. One change converts the cysteine (TGC) to phenylalanine (TTC) at amino acid 161 of the maize Protox-1 sequence (SEQ ID NO. 6). The second converts the isoleucine (ATA) to threonine (ACA) at amino acid 421.

EXAMPLE 8: Combining identified resistance mutations with identified second site mutations to create highly functional/highly tolerant protox enzymes

The AraC305Leu mutation described above was found to enhance the function/herbicide resistance of both the AraC-1Val and the AraC-2Cys mutant plasmids. In an effort to test the

general usefulness of this second site mutation, it was combined separately with the AraC-2Leu, AraC-2Val, and AraC-2Ile mutations and tested for herbicide tolerance. In each case, the AraC305Leu change significantly increased the growth rate of the resistant protox mutant on protox-inhibiting herbicide. Combinations of the AraC-2Ile resistant mutant with either the second site mutant AraC249Ile or AraC118Leu also produced more highly tolerant mutant protox enzymes. The AraC249Ile mutation demonstrates that a second site mutation identified as enhancing an AraC-1 mutant may also increase the resistance of an AraC-2 mutant. A three mutation plasmid containing AraC-2Ile, AraC305Leu, and AraC249Ile has also been shown to produce a highly functional, highly herbicide tolerant protox-1 enzyme.

EXAMPLE 9: Demonstration of resistant mutations' cross-tolerance to various protox-inhibiting compounds

Resistant mutant plasmids, originally identified based on resistance against a single protox inhibitory herbicide, were tested against a spectrum of other protox-inhibiting compounds. For this test, the SASX38 strain containing the wild-type plasmid is plated on a range of concentrations of each compound to determine the lethal concentration for each one. Resistant mutant plasmids in SASX38 are plated and scored for the ability to survive on a concentration of each compound which is at least 10 fold higher than the concentration that is lethal to the SASX38 strain containing the wild-type plasmid.

Results from cross-tolerance testing, illustrated in Tables 3A and 3B below, show that each of the mutations identified confer tolerance to a variety of protox inhibiting compounds.

Table 3A

Cross Tolerance of Plant Protox Mutants to Various Protox Inhibitors

	AmC-1Val	AmC-2Cys	AmC-1Thr	AmC-3Thr	MsC-1Val
CGA 276'854 - Ciba	+	+	+	+	+
CGA 248'757 - Kumiai	+	+	+	-	+
CGA 175'769 - Rohm-Haas	++	-	++	++	-
CGA 263'195 - Sumitomo	+	+	+	+	+
CGA 284'593 - Uniroyal	-	+	+	++	+
CGA 260670 - Sumitomo	-	-	-	-	+
CGA 333'855 - Nihon-Nohyaku	+	-	++	++	++
CGA 245'027 - Sumitomo	+	-	+	+	+
*CGA 302'640 - FMC					
*CGA 335'141 - FMC					

+ = 10X or more tolerant than WT

++ = 100X or more tolerant than WT

- = no cross tolerance

* = these compounds were tested but provided no information

Table 3B
Cross Tolerance of Plant Protox Mutants to Various Protox Inhibitors

	AmC-1Leu	AmC-2Ile	AmC-1Leu + AmC-2Met	AmC-1Leu + AmC-2Leu	AmC-2Ile + AmC305Leu	AmC-2Cys + AmC425Ser	AmC-2Leu + AmC425Ser	AmC-2Met + AmC425Ser
CGA 276'854 - Ciba	+	+	+	+	+	+	+	+
CGA 248'757 - Kumiai	++	++	++	++	++	++	++	++
Ciba 175'769 - Bohm-Haas	++	-	+	++	+	-	+	+
CGA 263'195 - Sumitomo	++	+++	+++	+++	+++	++	+++	++
CGA 284'593 - Uniroyal	++	++	++	++	++	++	++	++
CGA 260'670 - Sumitomo	+++	+++	+++	+++	+++	+	++	++
CGA 333'855 - Nihon-Nohyaku								
CGA 245'027 - Sumitomo	++	++	++	++	++	-	++	++

EXAMPLE 10: *Engineering of plants tolerant to protox-inhibiting herbicides by homologous recombination or gene conversion*

Because the described mutant coding sequences effectively confer herbicide tolerance when expressed under the control of the native protox promoter, targeted changes to the protox coding sequence in its native chromosomal location represent an alternative means for generating herbicide tolerant plants and plant cells. A fragment of protox DNA containing the desired mutations, but lacking its own expression signals (either promoter or 3' untranslated region) can be introduced by any of several art-recognized methods (for instance, Agrobacterium transformation, direct gene transfer to protoplasts, microprojectile bombardment), and herbicide-tolerant transformants selected. The introduced DNA fragment also contains a diagnostic restriction enzyme site or other sequence polymorphism that is introduced by site-directed mutagenesis *in vitro* without changing the encoded amino acid sequence (i.e. a silent mutation). As has been previously reported for various selectable marker and herbicide tolerance genes (see, *e.g.*, Paszkowski *et al.*, *EMBO J. 7:* 4021-4026 (1988); Lee *et al.*, *Plant Cell 2:* 415-425 (1990); Risseeuw *et al.*, *Plant J. 7:* 109-119 (1995)), some transformants are found to result from homologous integration of the mutant DNA into the protox chromosomal locus, or from conversion of the native protox chromosomal sequence to the introduced mutant sequence. These transformants are recognized by the combination of their herbicide-tolerant phenotype, and the presence of the diagnostic restriction enzyme site in their protox chromosomal locus.

EXAMPLE 11: Construction of Plant Transformation Vectors

Numerous transformation vectors are available for plant transformation, and the genes of this invention can be used in conjunction with any such vectors. The selection of vector for use will depend upon the preferred transformation technique and the target species for transformation. For certain target species, different antibiotic or herbicide selection markers may be preferred. Selection markers used routinely in transformation include the *nptII* gene which confers resistance to kanamycin and related antibiotics (Messing & Vierra, *Gene 19:* 259-268 (1982); Bevan *et al.*, *Nature 304:*184-187 (1983)), the *bar* gene which confers resistance to the herbicide

phosphinothricin (White *et al., Nucl Acids Res 18:* 1062 (1990), Spencer *et al.* Theor Appl Genet 79: 625-631(1990)), the *hph* gene which confers resistance to the antibiotic hygromycin (Blochinger & Diggelmann, *Mol Cell Biol 4:* 2929-2931), and the *dhfr* gene, which confers resistance to methotrexate (Bourouis *et al., EMBO J. 2(7):* 1099-1104 (1983)).

(1) Construction of Vectors Suitable for *Agrobacterium* Transformation

Many vectors are available for transformation using *Agrobacterium tumefaciens.* These typically carry at least one T-DNA border sequence and include vectors such as pBIN19 (Bevan, *Nucl. Acids Res.* (1984)) and pXYZ. Below the construction of two typical vectors is described.

Construction of pCIB200 and pCIB2001

The binary vectors pCIB200 and pCIB2001 are used for the construction of recombinant vectors for use with *Agrobacterium* and was constructed in the following manner. pTJS75kan was created by *NarI* digestion of pTJS75 (Schmidhauser & Helinski, *J Bacteriol. 164*: 446-455 (1985)) allowing excision of the tetracycline-resistance gene, followed by insertion of an *AccI* fragment from pUC4K carrying an NPTII (Messing & Vierra, *Gene 19*: 259-268 (1982); Bevan *et al., Nature 304*: 184-187 (1983); McBride *et al., Plant Molecular Biology 14*: 266-276 (1990)). *XhoI* linkers were ligated to the *EcoRV* fragment of pCIB7 which contains the left and right T-DNA borders, a plant selectable *nos/nptII* chimeric gene and the pUC polylinker (Rothstein *et al., Gene 53:* 153-161 (1987)), and the *XhoI*-digested fragment was cloned into *SalI*-digested pTJS75kan to create pCIB200 (see also EP 0 332 104, example 19). pCIB200 contains the following unique polylinker restriction sites: *EcoRI, SstI, KpnI, BglII, XbaI,* and *SalI. pCIB2001* is a derivative of pCIB200 which created by the insertion into the polylinker of additional restriction sites. Unique restriction sites in the polylinker of pCIB2001 are *EcoRI, SstI, KpnI, BglII, XbaI, SalI, MluI, BclI, AvrII, ApaI, HpaI,* and *StuI.* pCIB2001, in addition to containing these unique restriction sites also has plant and bacterial kanamycin selection, left and right T-

DNA borders for *Agrobacterium*-mediated transformation, the RK2-derived *trfA* function for mobilization between *E. coli* and other hosts, and the *OriT* and *OriV* functions also from RK2. The pCIB2001 polylinker is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

Construction of pCIB10 and Hygromycin Selection Derivatives thereof

The binary vector pCIB10 contains a gene encoding kanamycin resistance for selection in plants, T-DNA right and left border sequences and incorporates sequences from the wide host-range plasmid pRK252 allowing it to replicate in both *E. coli* and *Agrobacterium*. Its construction is described by Rothstein *et al.*, *Gene 53*: 153-161 (1987). Various derivatives of pCIB10 have been constructed which incorporate the gene for hygromycin B phosphotransferase described by Gritz *et al.*, *Gene 25*: 179-188 (1983)). These derivatives enable selection of transgenic plant cells on hygromycin only (pCIB743), or hygromycin and kanamycin (pCIB715, pCIB717).

(2) Construction of Vectors Suitable for non-*Agrobacterium* Transformation.

Transformation without the use of *Agrobacterium tumefaciens* circumvents the requirement for T-DNA sequences in the chosen transformation vector and consequently vectors lacking these sequences can be utilized in addition to vectors such as the ones described above which contain T-DNA sequences. Transformation techniques which do not rely on *Agrobacterium* include transformation via particle bombardment, protoplast uptake (*e.g.* PEG and electroporation) and microinjection. The choice of vector depends largely on the preferred selection for the species being transformed. Below, the construction of some typical vectors is described.

Construction of pCIB3064

pCIB3064 is a pUC-derived vector suitable for direct gene transfer techniques in combination with selection by the herbicide basta (or phosphinothricin). The plasmid pCIB246 comprises the CaMV 35S promoter in operational fusion to the *E. coli* GUS gene and the CaMV 35S transcriptional terminator and is described in the PCT published application WO 93/07278. The 35S promoter of this vector contains two ATG sequences 5' of the start site. These sites were mutated using standard PCR techniques in such a way as to remove the ATGs and generate the restriction sites *Sspl* and *PvuII*. The new restriction sites were 96 and 37 bp away from the unique *SalI* site and 101 and 42 bp away from the actual start site. The resultant derivative of pCIB246 was designated pCIB3025. The GUS gene was then excised from pCIB3025 by digestion with *SalI* and *SacI*, the termini rendered blunt and religated to generate plasmid pCIB3060. The plasmid pJIT82 was obtained from the John Innes Centre, Norwich and the a 400 bp *SmaI* fragment containing the *bar* gene from *Streptomyces viridochromogenes* was excised and inserted into the *HpaI* site of pCIB3060 (Thompson *et al.* EMBO J 6: 2519-2523 (1987)). This generated pCIB3064 which comprises the *bar* gene under the control of the CaMV 35S promoter and terminator for herbicide selection, a gene fro ampicillin resistance (for selection in

E. coli) and a polylinker with the unique sites *SphI, PstI, HindIII,* and *BamHI*. This vector is suitable for the cloning of plant expression cassettes containing their own regulatory signals.

Construction of pSOG19 and pSOG35

pSOG35 is a transformation vector which utilizes the *E. coli* gene dihydrofolate reductase (DHFR) as a selectable marker conferring resistance to methotrexate. PCR was used to amplify the 35S promoter (~800 bp), intron 6 from the maize Adh1 gene (~550 bp) and 18 bp of the GUS untranslated leader sequence from pSOG10. A 250 bp fragment encoding the *E. coli* dihydrofolate reductase type II gene was also amplified by PCR and these two PCR fragments were assembled with a *SacI-PstI* fragment from pBI221 (Clontech) which comprised the pUC19 vector backbone and the nopaline synthase terminator. Assembly of these fragments generated pSOG19 which contains the 35S promoter in fusion with the intron 6 sequence, the GUS leader, the DHFR gene and the nopaline synthase terminator. Replacement of the GUS leader in pSOG19 with the leader sequence from Maize Chlorotic Mottle Virus (MCMV) generated the vector pSOG35. pSOG19 and pSOG35 carry the pUC gene for ampicillin resistance and have *HindIII, SphI, PstI* and *EcoRI* sites available for the cloning of foreign sequences.

EXAMPLE 12: Construction of Plant Expression Cassettes

Gene sequences intended for expression in transgenic plants are firstly assembled in expression cassettes behind a suitable promoter and upstream of a suitable transcription terminator. These expression cassettes can then be easily transferred to the plant transformation vectors described above in Example 19.

Promoter Selection

The selection of a promoter used in expression cassettes will determine the spatial and temporal expression pattern of the transgene in the transgenic plant. Selected promoters will express transgenes in specific cell types (such as leaf epidermal cells, mesophyll cells, root cortex cells) or in specific tissues or organs (roots, leaves or flowers, for example) and this selection will

reflect the desired location of expression of the transgene. Alternatively, the selected promoter may drive expression of the gene under a light-induced or other temporally regulated promoter. A further alternative is that the selected promoter be chemically regulated. This would provide the possibility of inducing expression of the transgene only when desired and caused by treatment with a chemical inducer.

Transcriptional Terminators

A variety of transcriptional terminators are available for use in expression cassettes. These are responsible for the termination of transcription beyond the transgene and its correct polyadenylation. Appropriate transcriptional terminators are those which are known to function in plants and include the CaMV 35S terminator, the *tml* terminator, the nopaline synthase terminator, the pea *rbcS* E9 terminator, as well as terminators naturally associated with the plant protox gene (i.e. "ptotox terminators"). These can be used in both monocotyledons and dicotyledons.

Sequences for the Enhancement or Regulation of Expression

Numerous sequences have been found to enhance gene expression from within the transcriptional unit and these sequences can be used in conjunction with the genes of this invention to increase their expression in transgenic plants.

Various intron sequences have been shown to enhance expression, particularly in monocotyledonous cells. For example, the introns of the maize *Adh1* gene have been found to significantly enhance the expression of the wild-type gene under its cognate promoter when introduced into maize cells. Intron 1 was found to be particularly effective and enhanced expression in fusion constructs with the chloramphenicol acetyltransferase gene (Callis *et al.*, Genes Develop. 1: 1183-1200 (1987)). In the same experimental system, the intron from the maize *bronze1* gene had a similar effect in enhancing expression (Callis *et al.*, *supra*). Intron sequences have been routinely incorporated into plant transformation vectors, typically within the non-translated leader.

A number of non-translated leader sequences derived from viruses are also known to enhance expression, and these are particularly effective in dicotyledonous cells. Specifically,

leader sequences from Tobacco Mosaic Virus (TMV, the "W-sequence"), Maize Chlorotic Mottle Virus (MCMV), and Alfalfa Mosaic Virus (AMV) have been shown to be effective in enhancing expression (*e.g.* Gallie *et al. Nucl. Acids Res. 15*: 8693-8711 (1987); Skuzeski *et al.* Plant Molec. Biol. 15: 65-79 (1990))

Targeting of the Gene Product Within the Cell

Various mechanisms for targeting gene products are known to exist in plants and the sequences controlling the functioning of these mechanisms have been characterized in some detail. For example, the targeting of gene products to the chloroplast is controlled by a signal sequence found at the amino terminal end of various proteins and which is cleaved during chloroplast import yielding the mature protein (*e.g.* Comai *et al. J. Biol. Chem. 263*: 15104-15109 (1988)). These signal sequences can be fused to heterologous gene products to effect the import of heterologous products into the chloroplast (van den Broeck *et al. Nature 313*: 358-363 (1985)). DNA encoding for appropriate signal sequences can be isolated from the 5' end of the cDNAs encoding the RUBISCO protein, the CAB protein, the EPSP synthase enzyme, the GS2 protein and many other proteins which are known to be chloroplast localized.

Other gene products are localized to other organelles such as the mitochondrion and the peroxisome (*e.g.* Unger *et al. Plant Molec. Biol. 13*: 411-418 (1989)). The cDNAs encoding these products can also be manipulated to effect the targeting of heterologous gene products to these organelles. Examples of such sequences are the nuclear-encoded ATPases and specific aspartate amino transferase isoforms for mitochondria. Targeting to cellular protein bodies has been described by Rogers *et al., Proc. Natl. Acad. Sci. USA* 82: 6512-6516 (1985)).

In addition sequences have been characterized which cause the targeting of gene products to other cell compartments. Amino terminal sequences are responsible for targeting to the ER, the apoplast, and extracellular secretion from aleurone cells (Koehler & Ho, *Plant Cell* 2: 769-783 (1990)). Additionally, amino terminal sequences in conjunction with carboxy terminal sequences are responsible for vacuolar targeting of gene products (Shinshi *et al., Plant Molec. Biol. 14*: 357-368 (1990)).

By the fusion of the appropriate targeting sequences described above to transgene sequences of interest it is possible to direct the transgene product to any organelle or cell

compartment. For chloroplast targeting, for example, the chloroplast signal sequence from the RUBISCO gene, the CAB gene, the EPSP synthase gene, or the GS2 gene is fused in frame to the amino terminal ATG of the transgene. The signal sequence selected should include the known cleavage site and the fusion constructed should take into account any amino acids after the cleavage site which are required for cleavage. In some cases this requirement may be fulfilled by the addition of a small number of amino acids between the cleavage site and the transgene ATG or alternatively replacement of some amino acids within the transgene sequence. Fusions constructed for chloroplast import can be tested for efficacy of chloroplast uptake by *in vitro* translation of *in vitro* transcribed constructions followed by *in vitro* chloroplast uptake using techniques described by (Bartlett *et al.* In: Edelmann *et al.* (Eds.) Methods in Chloroplast Molecular Biology, Elsevier. pp 1081-1091 (1982); Wasmann *et al. Mol. Gen. Genet. 205*: 446-453 (1986)). These construction techniques are well known in the art and are equally applicable to mitochondria and peroxisomes. The choice of targeting which may be required for expression of the transgenes will depend on the cellular localization of the precursor required as the starting point for a given pathway. This will usually be cytosolic or chloroplastic, although it may is some cases be mitochondrial or peroxisomal. The products of transgene expression will not normally require targeting to the ER, the apoplast or the vacuole.

The above described mechanisms for cellular targeting can be utilized not only in conjunction with their cognate promoters, but also in conjunction with heterologous promoters so as to effect a specific cell targeting goal under the transcriptional regulation of a promoter which has an expression pattern different to that of the promoter from which the targeting signal derives.

EXAMPLE 13: Transformation of Dicotyledons

Transformation techniques for dicotyledons are well known in the art and include *Agrobacterium*-based techniques and techniques which do not require *Agrobacterium*. Non-*Agrobacterium* techniques involve the uptake of exogenous genetic material directly by protoplasts or cells. This can be accomplished by PEG or electroporation mediated uptake, particle bombardment-mediated delivery, or microinjection. Examples of these techniques are described by Paszkowski *et al.*, *EMBO J 3:* 2717-2722 (1984), Potrykus *et al.*, *Mol. Gen. Genet. 199:* 169-177 (1985), Reich *et al.*, *Biotechnology 4:* 1001-1004 (1986), and Klein *et al.*, *Nature*

327: 70-73 (1987). In each case the transformed cells are regenerated to whole plants using standard techniques known in the art.

Agrobacterium-mediated transformation is a preferred technique for transformation of dicotyledons because of its high efficiency of transformation and its broad utility with many different species. The many crop species which are routinely transformable by *Agrobacterium* include tobacco, tomato, sunflower, cotton, oilseed rape, potato, soybean, alfalfa and poplar (EP 0 317 511 (cotton), EP 0 249 432 (tomato, to Calgene), WO 87/07299 (*Brassica*, to Calgene), US 4,795,855 (poplar)).

Transformation of the target plant species by recombinant *Agrobacterium* usually involves co-cultivation of the *Agrobacterium* with explants from the plant and follows protocols well known in the art. Transformed tissue is regenerated on selectable medium carrying the antibiotic or herbicide resistance marker present between the binary plasmid T-DNA borders.

EXAMPLE 14: Transformation of Monocotyledons

Transformation of most monocotyledon species has now also become routine. Preferred techniques include direct gene transfer into protoplasts using PEG or electroporation techniques, and particle bombardment into callus tissue. Transformations can be undertaken with a single DNA species or multiple DNA species (*i.e.* co-transformation) and both these techniques are suitable for use with this invention. Co-transformation may have the advantage of avoiding complex vector construction and of generating transgenic plants with unlinked loci for the gene of interest and the selectable marker, enabling the removal of the selectable marker in subsequent generations, should this be regarded desirable. However, a disadvantage of the use of co-transformation is the less than 100% frequency with which separate DNA species are integrated into the genome (Schocher *et al. Biotechnology 4:* 1093-1096 (1986)).

Patent Applications EP 0 292 435 (to Ciba-Geigy), EP 0 392 225 (to Ciba-Geigy) and WO 93/07278 (to Ciba-Geigy) describe techniques for the preparation of callus and protoplasts from an élite inbred line of maize, transformation of protoplasts using PEG or electroporation, and the regeneration of maize plants from transformed protoplasts. Gordon-Kamm *et al.*, *Plant Cell 2:* 603-618 (1990)) and Fromm *et al.*, *Biotechnology 8:* 833-839 (1990)) have published techniques for transformation of A188-derived maize line using particle bombardment.

Furthermore, application WO 93/07278 (to Ciba-Geigy) and Koziel *et al.*, *Biotechnology 11:* 194-200 (1993)) describe techniques for the transformation of elite inbred lines of maize by particle bombardment. This technique utilizes immature maize embryos of 1.5-2.5 mm length excised from a maize ear 14-15 days after pollination and a PDS-1000He Biolistics device for bombardment.

Transformation of rice can also be undertaken by direct gene transfer techniques utilizing protoplasts or particle bombardment. Protoplast-mediated transformation has been described for *Japonica*-types and *Indica*-types (Zhang *et al.*, *Plant Cell Rep 7*: 379-384 (1988); Shimamoto *et al. Nature 338*: 274-277 (1989); Datta *et al. Biotechnology 8:* 736-740 (1990)). Both types are also routinely transformable using particle bombardment (Christou *et al. Biotechnology 9:* 957-962 (1991)).

Patent Application EP 0 332 581 (to Ciba-Geigy) describes techniques for the generation, transformation and regeneration of Pooideae protoplasts. These techniques allow the transformation of *Dactylis* and wheat. Furthermore, wheat transformation was been described by Vasil *et al.*, *Biotechnology 10:* 667-674 (1992)) using particle bombardment into cells of type C long-term regenerable callus, and also by Vasil *et al.*, *Biotechnology 11:* 1553-1558 (1993)) and Weeks *et al.*, *Plant Physiol. 102*: 1077-1084 (1993) using particle bombardment of immature embryos and immature embryo-derived callus. A preferred technique for wheat transformation, however, involves the transformation of wheat by particle bombardment of immature embryos and includes either a high sucrose or a high maltose step prior to gene delivery. Prior to bombardment, any number of embryos (0.75-1 mm in length) are plated onto MS medium with 3% sucrose (Murashige & Skoog, *Physiologia Plantarum 15*: 473-497 (1962)) and 3 mg/l 2,4-D for induction of somatic embryos which is allowed to proceed in the dark. On the chosen day of bombardment, embryos are removed from the induction medium and placed onto the osmoticum (*i.e.* induction medium with sucrose or maltose added at the desired concentration, typically 15%). The embryos are allowed to plasmolyze for 2-3 h and are then bombarded. Twenty embryos per target plate is typical, although not critical. An appropriate gene-carrying plasmid (such as pCIB3064 or pSG35) is precipitated onto micrometer size gold particles using standard procedures. Each plate of embryos is shot with the DuPont Biolistics· helium device using a burst pressure of ~1000 psi using a standard 80 mesh screen. After bombardment, the embryos are

aced back into the dark to recover for about 24 h (still on osmoticum). After 24 hrs, the embryos are removed from the osmoticum and placed back onto induction medium where they stay for about a month before regeneration. Approximately one month later the embryo explants with developing embryogenic callus are transferred to regeneration medium (MS + 1 mg/liter NAA, 5 mg/liter GA), further containing the appropriate selection agent (10 mg/l basta in the case of pCIB3064 and 2 mg/l methotrexate in the case of pSOG35). After approximately one month, developed shoots are transferred to larger sterile containers known as "GA7s" which contained half-strength MS, 2% sucrose, and the same concentration of selection agent. Patent application 08/147,161 describes methods for wheat transformation and is hereby incorporated by reference.

EXAMPLE 15: Isolation of the Arabidopsis thaliana Protox-1 promoter sequence

A Lambda Zap II genomic DNA library prepared from *Arabidopsis thaliana* (Columbia, whole plant) was purchased from Stratagene. Approximately 125,000 phage were plated at a density of 25,000 pfu per 15 cm Petri dish and duplicate lifts were made onto Colony/Plaque Screen membranes (NEN Dupont). The plaque lifts were probed with the Arabidopsis Protox-1 cDNA (SEQ ID No. 1 labeled with 32P-dCTP by the random priming method (Life Technologies). Hybridization and wash conditions were at 65° C as described in Church and Gilbert, *Proc. Natl. Acad. Sci. USA 81:* 1991-1995 (1984). Positively hybridizing plaques were purified and in vivo excised into pBluescript plasmids. Sequence from the genomic DNA inserts was determined by the chain termination method using dideoxy terminators labeled with fluorescent dyes (Applied Biosystems, Inc.). One clone, AraPT1Pro, was determined to contain 580 bp of Arabidopsis sequence upstream from the initiating methionine (ATG) of the Protox-1 protein coding sequence. This clone also contains coding sequence and introns that extend to bp 1241 of the Protox-1 cDNA sequence. The 580 bp 5' noncoding fragment is the putative Arabidopsis Protox-1 promoter, and the sequence is set forth in SEQ ID No. 13.

AraPT1Pro was deposited December 14, 1995, as pWDC-11 (NRRL #B-21515)

EXAMPLE 16: Construction of plant transformation vectors expressing altered Protox-1 genes behind the native Arabidopsis Protox-1 promoter

A full-length cDNA of the appropriate altered *Arabidopsis* Protox-1 cDNA is isolated as an *EcoRI*-*XhoI* partial digest fragment and cloned into the plant expression vector pCGN1761ENX (see Example 9 of International application no. PCT/IB95/00452 filed June 8, 1995, published Dec. 21, 1995 as WO 95/34659). This plasmid is digested with *NcoI* and *BamHI* to produce a fragment comprised of the complete Protox-1 cDNA plus a transcription terminator from the 3' untranslated sequence of the tml gene of *Agrobacterium tumefaciens*. The AraPT1Pro plasmid described above is digested with NcoI and BamHI to produce a fragment comprised of pBluescript and the 580 bp putative Arabidopsis Protox-1 promoter. Ligation of these two fragments produces a fusion of the altered protox cDNA to the native protox promoter. The expression cassette containing the Protox-1 promoter/Protox-1 cDNA/tml terminator fusion is excised by digestion with KpnI and cloned into the binary vector pCIB200. The binary plasmid is transformed by electroporation into Agrobacterium and then into Arabidopsis using the vacuum infiltration method (Bechtold *et al. C.R. Acad. Sci. Paris 316:* 1194-1199 (1993). Transformants expressing altered protox genes are selected on kanamycin or on various concentrations of protox inhibiting herbicide.

EXAMPLE 17: Production of herbicide tolerant plants by expression of a native Protox-1 promoter/altered Prctox-1 fusion

Using the procedure described above, an *Arabidopsis* Protox-1 cDNA containing a TAC to ATG (Tyrosine to Methionine) change at nucleotides 1306-1308 in the Protox-1 sequence (SEQ ID No.1) was fused to the native Protox-1 promoter fragment and transformed into *Arabidopsis thaliana*. This altered Protox-1 enzyme (AraC-2Met) has been shown to be >10fold more tolerant to various protox-inhibiting herbicides than the naturally occurring enzyme when tested in a bacterial expression system (see Examples 5-9). Seed from the vacuum infiltrated plants was collected and plated on a range (10.0nM-1.0uM) of a protox inhibitory aryluracil herbicide of formula XVII. Multiple experiments with wild type *Arabidopsis* have shown that a 10.0nM concentration of this compound is sufficient to prevent normal seedling germination. Transgenic seeds expressing the AraC-2Met altered enzyme fused to the native Protox-1

plant produced normal Arabidopsis seedlings at herbicide concentrations up to 500nM, indicating at least 50-fold higher herbicide tolerance when compared to wild-type Arabidopsis. This promoter/altered protox enzyme fusion therefore functions as an effective selectable marker for plant transformation. Several of the plants that germinated on 100.0uM of protox-inhibiting herbicide were transplanted to soil, grown 2-3 weeks, and tested in a spray assay with various concentrations of the protox-inhibiting herbicide. When compared to empty vector control transformants, the AraPT1Pro/AraC-2Met transgenics were >10fold more tolerant to the herbicide spray.

Various modifications of the invention described herein will become apparent to those skilled in the art. Such modifications are intended to fall within the scope of the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Volrath, Sandra L.
Pacella, Marie A.
Potter, Sharon L.
Ward, Eric R.

(ii) TITLE OF INVENTION: DNA MOLECULES ENCODING PLANT PROTOPORPHYRINOGEN OXIDASE AND INHIBITOR-RESISTANT MUTANTS THEREOF

(iii) NUMBER OF SEQUENCES: 13

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Ciba-Geigy Corporation / Patent Dept.
(B) STREET: 540 White Plains Rd.
(C) CITY: Tarrytown
(D) STATE: NY
(E) COUNTRY: USA
(F) ZIP: 10591-9005

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US TBA
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA
(A) APPLICATION NUMBER: US 08/261,198
(B) FILING DATE: 16-JUN-94

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Elmer, James Scott
(B) REGISTRATION NUMBER: 36,129
(C) REFERENCE/DOCKET NUMBER: CGC 1847/prov

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 919-541-8614
(B) TELEFAX: 919-541-8689

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1719 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 31..1644
(D) OTHER INFORMATION: /note= "Arabidopsis protox-1 cDNA; sequence from pNDC-2"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
TGACAAAATT CCGAATTCTC TGCGATTTCC ATG GAG TTA TCT CTT CTC CGT CCG        54
                                 Met Glu Leu Ser Leu Leu Arg Pro
                                   1               5

ACG ACT CAA TCG CTT CTT CCG TCG TTT TCG AAG CCC AAT CTC CGA TTA       102
Thr Thr Gln Ser Leu Leu Pro Ser Phe Ser Lys Pro Asn Leu Arg Leu
     10                  15                  20

AAT GTT TAT AAG CCT CTT AGA CTC CGT TGT TCA GTG GCC GGT GGA CCA       150
Asn Val Tyr Lys Pro Leu Arg Leu Arg Cys Ser Val Ala Gly Gly Pro
 25                  30                  35                  40

ACC GTC GGA TCT TCA AAA ATC GAA GGC GGA GGA GGC ACC ACC ATC ACG       198
Thr Val Gly Ser Ser Lys Ile Glu Gly Gly Gly Gly Thr Thr Ile Thr
                 45                  50                  55

ACG GAT TGT GTC ATT GTC GGC GGA GGT ATT AGT GGT CTT TGC ATC GCT       246
Thr Asp Cys Val Ile Val Gly Gly Gly Ile Ser Gly Leu Cys Ile Ala
             60                  65                  70

CAG GCG CTT GCT ACT AAG CAT CCT GAT GCT GCT CCG AAT TTA ATT GTG       294
Gln Ala Leu Ala Thr Lys His Pro Asp Ala Ala Pro Asn Leu Ile Val
         75                  80                  85

ACC GAG GCT AAG GAT CGT GTT GGA GGC AAC ATT ATC ACT CGT GAA GAG       342
Thr Glu Ala Lys Asp Arg Val Gly Gly Asn Ile Ile Thr Arg Glu Glu
     90                  95                 100

AAT GGT TTT CTC TGG GAA GAA GGT CCC AAT AGT TTT CAA CCG TCT GAT       390
Asn Gly Phe Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp
105                 110                 115                 120

CCT ATG CTC ACT ATG GTG GTA GAT AGT GGT TTG AAG GAT GAT TTG GTG       438
Pro Met Leu Thr Met Val Val Asp Ser Gly Leu Lys Asp Asp Leu Val
                125                 130                 135

TTG GGA GAT CCT ACT GCG CCA AGG TTT GTG TTG TGG AAT GGG AAA TTG       486
Leu Gly Asp Pro Thr Ala Pro Arg Phe Val Leu Trp Asn Gly Lys Leu
            140                 145                 150

AGG CCG GTT CCA TCG AAG CTA ACA GAC TTA CCG TTC TTT GAT TTG ATG       534
Arg Pro Val Pro Ser Lys Leu Thr Asp Leu Pro Phe Phe Asp Leu Met
        155                 160                 165

AGT ATT GGT GGG AAG ATT AGA GCT GGT TTT GGT GCA CTT GGC ATT CGA       582
Ser Ile Gly Gly Lys Ile Arg Ala Gly Phe Gly Ala Leu Gly Ile Arg
    170                 175                 180

CCG TCA CCT CCA GGT CGT GAA GAA TCT GTG GAG GAG TTT GTA CGG CGT       630
Pro Ser Pro Pro Gly Arg Glu Glu Ser Val Glu Glu Phe Val Arg Arg
185                 190                 195                 200

AAC CTC GGT GAT GAG GTT TTT GAG CGC CTG ATT GAA CCG TTT TGT TCA       678
Asn Leu Gly Asp Glu Val Phe Glu Arg Leu Ile Glu Pro Phe Cys Ser
```

```
                    205                 210                 215

GGT GTT TAT GCT GGT GAT CCT TCA AAA CTG AGC ATG AAA GCA GCG TTT   726
Gly Val Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala Ala Phe
            220                 225                 230

GGG AAG GTT TGG AAA CTA GAG CAA AAT GGT GGA AGC ATA ATA GGT GGT   774
Gly Lys Val Trp Lys Leu Glu Gln Asn Gly Gly Ser Ile Ile Gly Gly
        235                 240                 245

ACT TTT AAG GCA ATT CAG GAG AGG AAA AAC GCT CCC AAG GCA GAA CGA   822
Thr Phe Lys Ala Ile Gln Glu Arg Lys Asn Ala Pro Lys Ala Glu Arg
    250                 255                 260

GAC CCG CGC CTG CCA AAA CCA CAG GGC CAA ACA GTT GGT TCT TTC AGG   870
Asp Pro Arg Leu Pro Lys Pro Gln Gly Gln Thr Val Gly Ser Phe Arg
265                 270                 275                 280

AAG GGA CTT CGA ATG TTG CCA GAA GCA ATA TCT GCA AGA TTA GGT AGC   918
Lys Gly Leu Arg Met Leu Pro Glu Ala Ile Ser Ala Arg Leu Gly Ser
                285                 290                 295

AAA GTT AAG TTG TCT TGG AAG CTC TCA GGT ATC ACT AAG CTG GAG AGC   966
Lys Val Lys Leu Ser Trp Lys Leu Ser Gly Ile Thr Lys Leu Glu Ser
            300                 305                 310

GGA GGA TAC AAC TTA ACA TAT GAG ACT CCA GAT GGT TTA GTT TCC GTG  1014
Gly Gly Tyr Asn Leu Thr Tyr Glu Thr Pro Asp Gly Leu Val Ser Val
        315                 320                 325

CAG AGC AAA AGT GTT GTA ATG ACG GTG CCA TCT CAT GTT GCA AGT GGT  1062
Gln Ser Lys Ser Val Val Met Thr Val Pro Ser His Val Ala Ser Gly
    330                 335                 340

CTC TTG CGC CCT CTT TCT GAA TCT GCT GCA AAT GCA CTC TCA AAA CTA  1110
Leu Leu Arg Pro Leu Ser Glu Ser Ala Ala Asn Ala Leu Ser Lys Leu
345                 350                 355                 360

TAT TAC CCA CCA GTT GCA GCA GTA TCT ATC TCG TAC CCG AAA GAA GCA  1158
Tyr Tyr Pro Pro Val Ala Ala Val Ser Ile Ser Tyr Pro Lys Glu Ala
                365                 370                 375

ATC CGA ACA GAA TGT TTG ATA GAT GGT GAA CTA AAG GGT TTT GGG CAA  1206
Ile Arg Thr Glu Cys Leu Ile Asp Gly Glu Leu Lys Gly Phe Gly Gln
            380                 385                 390

TTG CAT CCA CGC ACG CAA GGA GTT GAA ACA TTA GGA ACT ATC TAC AGC  1254
Leu His Pro Arg Thr Gln Gly Val Glu Thr Leu Gly Thr Ile Tyr Ser
        395                 400                 405

TCC TCA CTC TTT CCA AAT CGC GCA CCG CCC GGA AGA ATT TTG CTG TTG  1302
Ser Ser Leu Phe Pro Asn Arg Ala Pro Pro Gly Arg Ile Leu Leu Leu
    410                 415                 420

AAC TAC ATT GGC GGG TCT ACA AAC ACC GGA ATT CTG TCC AAG TCT GAA  1350
Asn Tyr Ile Gly Gly Ser Thr Asn Thr Gly Ile Leu Ser Lys Ser Glu
425                 430                 435                 440

GGT GAG TTA GTG GAA GCA GTT GAC AGA GAT TTG AGG AAA ATG CTA ATT  1398
Gly Glu Leu Val Glu Ala Val Asp Arg Asp Leu Arg Lys Met Leu Ile
                445                 450                 455

AAG CCT AAT TCG ACC GAT CCA CTT AAA TTA GGA GTT AGG GTA TGG CCT  1446
Lys Pro Asn Ser Thr Asp Pro Leu Lys Leu Gly Val Arg Val Trp Pro
```

```
                460                 465                 470

CAA GCC ATT CCT CAG TTT CTA GTT GGT CAC TTT GAT ATC CTT GAC ACG      1494
Gln Ala Ile Pro Gln Phe Leu Val Gly His Phe Asp Ile Leu Asp Thr
        475                 480                 485

GCT AAA TCA TCT CTA ACG TCT TCG GGC TAC GAA GGG CTA TTT TTG GGT      1542
Ala Lys Ser Ser Leu Thr Ser Ser Gly Tyr Glu Gly Leu Phe Leu Gly
    490                 495                 500

GGC AAT TAC GTC GCT GGT GTA GCC TTA GGC CGG TGT GTA GAA GGC GCA      1590
Gly Asn Tyr Val Ala Gly Val Ala Leu Gly Arg Cys Val Glu Gly Ala
505                 510                 515                 520

TAT GAA ACC GCG ATT GAG GTC AAC AAC TTC ATG TCA CGC TAC GCT TAC      1638
Tyr Glu Thr Ala Ile Glu Val Asn Asn Phe Met Ser Arg Tyr Ala Tyr
                525                 530                 535

AAG TAAATGTAAA ACATTAAATC TCCCAGCTTG CGTGAGTTTT ATTAAATATT           1691
Lys

TTGAGATATC CAAAAAAAAA AAAAAAAA                                       1719
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 537 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Glu Leu Ser Leu Leu Arg Pro Thr Thr Gln Ser Leu Leu Pro Ser
  1               5                  10                  15

Phe Ser Lys Pro Asn Leu Arg Leu Asn Val Tyr Lys Pro Leu Arg Leu
            20                  25                  30

Arg Cys Ser Val Ala Gly Gly Pro Thr Val Gly Ser Ser Lys Ile Glu
        35                  40                  45

Gly Gly Gly Gly Thr Thr Ile Thr Thr Asp Cys Val Ile Val Gly Gly
    50                  55                  60

Gly Ile Ser Gly Leu Cys Ile Ala Gln Ala Leu Ala Thr Lys His Pro
65                  70                  75                  80

Asp Ala Ala Pro Asn Leu Ile Val Thr Glu Ala Lys Asp Arg Val Gly
                85                  90                  95

Gly Asn Ile Ile Thr Arg Glu Glu Asn Gly Phe Leu Trp Glu Glu Gly
            100                 105                 110

Pro Asn Ser Phe Gln Pro Ser Asp Pro Met Leu Thr Met Val Val Asp
        115                 120                 125

Ser Gly Leu Lys Asp Asp Leu Val Leu Gly Asp Pro Thr Ala Pro Arg
    130                 135                 140

Phe Val Leu Trp Asn Gly Lys Leu Arg Pro Val Pro Ser Lys Leu Thr
```

145 150 155 160

Asp Leu Pro Phe Phe Asp Leu Met Ser Ile Gly Gly Lys Ile Arg Ala
165 170 175

Gly Phe Gly Ala Leu Gly Ile Arg Pro Ser Pro Pro Gly Arg Glu Glu
180 185 190

Ser Val Glu Glu Phe Val Arg Arg Asn Leu Gly Asp Glu Val Phe Glu
195 200 205

Arg Leu Ile Glu Pro Phe Cys Ser Gly Val Tyr Ala Gly Asp Pro Ser
210 215 220

Lys Leu Ser Met Lys Ala Ala Phe Gly Lys Val Trp Lys Leu Glu Gln
225 230 235 240

Asn Gly Gly Ser Ile Ile Gly Gly Thr Phe Lys Ala Ile Gln Glu Arg
245 250 255

Lys Asn Ala Pro Lys Ala Glu Arg Asp Pro Arg Leu Pro Lys Pro Gln
260 265 270

Gly Gln Thr Val Gly Ser Phe Arg Lys Gly Leu Arg Met Leu Pro Glu
275 280 285

Ala Ile Ser Ala Arg Leu Gly Ser Lys Val Lys Leu Ser Trp Lys Leu
290 295 300

Ser Gly Ile Thr Lys Leu Glu Ser Gly Gly Tyr Asn Leu Thr Tyr Glu
305 310 315 320

Thr Pro Asp Gly Leu Val Ser Val Gln Ser Lys Ser Val Val Met Thr
325 330 335

Val Pro Ser His Val Ala Ser Gly Leu Leu Arg Pro Leu Ser Glu Ser
340 345 350

Ala Ala Asn Ala Leu Ser Lys Leu Tyr Tyr Pro Pro Val Ala Ala Val
355 360 365

Ser Ile Ser Tyr Pro Lys Glu Ala Ile Arg Thr Glu Cys Leu Ile Asp
370 375 380

Gly Glu Leu Lys Gly Phe Gly Gln Leu His Pro Arg Thr Gln Gly Val
385 390 395 400

Glu Thr Leu Gly Thr Ile Tyr Ser Ser Ser Leu Phe Pro Asn Arg Ala
405 410 415

Pro Pro Gly Arg Ile Leu Leu Leu Asn Tyr Ile Gly Gly Ser Thr Asn
420 425 430

Thr Gly Ile Leu Ser Lys Ser Glu Gly Glu Leu Val Glu Ala Val Asp
435 440 445

Arg Asp Leu Arg Lys Met Leu Ile Lys Pro Asn Ser Thr Asp Pro Leu
450 455 460

Lys Leu Gly Val Arg Val Trp Pro Gln Ala Ile Pro Gln Phe Leu Val
465 470 475 480

Gly His Phe Asp Ile Leu Asp Thr Ala Lys Ser Ser Leu Thr Ser Ser
485 490 495

```
Gly Tyr Glu Gly Leu Phe Leu Gly Gly Asn Tyr Val Ala Gly Val Ala
            500                 505                 510

Leu Gly Arg Cys Val Glu Gly Ala Tyr Glu Thr Ala Ile Glu Val Asn
        515                 520                 525

Asn Phe Met Ser Arg Tyr Ala Tyr Lys
    530                 535
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1738 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 70..1596
(D) OTHER INFORMATION: /note= 'Arabidopsis protox-2 cDNA; sequence from pMDC-1'

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
TTTTTTACTT ATTTCCCTCA CTCCTTTCGA CTGGTCAGAG ATTTTGACTC TGAATTGTTG      60

CAGATAGCA ATG GCG TCT GGA GCA GTA GCA GAT CAT CAA ATT GAA GCG          108
          Met Ala Ser Gly Ala Val Ala Asp His Gln Ile Glu Ala
            1               5                  10

GTT TCA GGA AAA AGA GTC GCA GTC GTA GGT GCA GGT GTA AGT GGA CTT        156
Val Ser Gly Lys Arg Val Ala Val Val Gly Ala Gly Val Ser Gly Leu
     15                  20                  25

GCG GCG GCT TAC AAG TTG AAA TCG AGG GGT TTG AAT GTC ACT GTC TTT        204
Ala Ala Ala Tyr Lys Leu Lys Ser Arg Gly Leu Asn Val Thr Val Phe
 30                  35                  40                  45

GAA GCT GAT GGA AGA GTA GGT GGG AAG TTG AGA AGT GTT ATG CAA AAT        252
Glu Ala Asp Gly Arg Val Gly Gly Lys Leu Arg Ser Val Met Gln Asn
                 50                  55                  60

GGT TTG ATT TGG GAT GAA GGA GCA AAC ACC ATG ACT GAG GCT GAG CCA        300
Gly Leu Ile Trp Asp Glu Gly Ala Asn Thr Met Thr Glu Ala Glu Pro
             65                  70                  75

GAA GTT GGG AGT TTA CTT GAT GAT CTT GGG CTT CGT GAG AAA CAA CAA        348
Glu Val Gly Ser Leu Leu Asp Asp Leu Gly Leu Arg Glu Lys Gln Gln
         80                  85                  90

TTT CCA ATT TCA CAG AAA AAG CGG TAT ATT GTG CGG AAT GGT GTA CCT        396
Phe Pro Ile Ser Gln Lys Lys Arg Tyr Ile Val Arg Asn Gly Val Pro
     95                  100                 105
```

```
GTG ATG CTA CCT ACC AAT CCC ATA GAG CTG GTC ACA AGT AGT GTG CTC          444
Val Met Leu Pro Thr Asn Pro Il  Glu Leu Val Thr Ser Ser Val Leu
110                 115                 120                 125

TCT ACC CAA TCT AAG TTT CAA ATC TTG TTG GAA CCA TTT TTA TGG AAG          492
Ser Thr Gln Ser Lys Phe Gln Ile Leu Leu Glu Pro Phe Leu Trp Lys
                    130                 135                 140

AAA AAG TCC TCA AAA GTC TCA GAT GCA TCT GCT GAA GAA AGT GTA AGC          540
Lys Lys Ser Ser Lys Val Ser Asp Ala Ser Ala Glu Glu Ser Val Ser
                145                 150                 155

GAG TTC TTT CAA CGC CAT TTT GGA CAA GAG GTT GTT GAC TAT CTC ATC          588
Glu Phe Phe Gln Arg His Phe Gly Gln Glu Val Val Asp Tyr Leu Ile
            160                 165                 170

GAC CCT TTT GTT GGT GGA ACA AGT GCT GCG GAC CCT GAT TCC CTT TCA          636
Asp Pro Phe Val Gly Gly Thr Ser Ala Ala Asp Pro Asp Ser Leu Ser
        175                 180                 185

ATG AAG CAT TCT TTC CCA GAT CTC TGG AAT GTA GAG AAA AGT TTT GGC          684
Met Lys His Ser Phe Pro Asp Leu Trp Asn Val Glu Lys Ser Phe Gly
190                 195                 200                 205

TCT ATT ATA GTC GGT GCA ATC AGA ACA AAG TTT GCT GCT AAA GGT GGT          732
Ser Ile Ile Val Gly Ala Ile Arg Thr Lys Phe Ala Ala Lys Gly Gly
                210                 215                 220

AAA AGT AGA GAC ACA AAG AGT TCT CCT GGC ACA AAA AAG GGT TCG CGT          780
Lys Ser Arg Asp Thr Lys Ser Ser Pro Gly Thr Lys Lys Gly Ser Arg
            225                 230                 235

GGG TCA TTC TCT TTT AAG GGG GGA ATG CAG ATT CTT CCT GAT ACG TTG          828
Gly Ser Phe Ser Phe Lys Gly Gly Met Gln Ile Leu Pro Asp Thr Leu
        240                 245                 250

TGC AAA AGT CTC TCA CAT GAT GAG ATC AAT TTA GAC TCC AAG GTA CTC          876
Cys Lys Ser Leu Ser His Asp Glu Ile Asn Leu Asp Ser Lys Val Leu
    255                 260                 265

TCT TTG TCT TAC AAT TCT GGA TCA AGA CAG GAG AAC TGG TCA TTA TCT          924
Ser Leu Ser Tyr Asn Ser Gly Ser Arg Gln Glu Asn Trp Ser Leu Ser
270                 275                 280                 285

TGT GTT TCG CAT AAT GAA ACG CAG AGA CAA AAC CCC CAT TAT GAT GCT          972
Cys Val Ser His Asn Glu Thr Gln Arg Gln Asn Pro His Tyr Asp Ala
                290                 295                 300

GTA ATT ATG ACG GCT CCT CTG TGC AAT GTG AAG GAG ATG AAG GTT ATG         1020
Val Ile Met Thr Ala Pro Leu Cys Asn Val Lys Glu Met Lys Val Met
            305                 310                 315

AAA GGA GGA CAA CCC TTT CAG CTA AAC TTT CTC CCC GAG ATT AAT TAC         1068
Lys Gly Gly Gln Pro Phe Gln Leu Asn Phe Leu Pro Glu Ile Asn Tyr
        320                 325                 330

ATG CCC CTC TCG GTT TTA ATC ACC ACA TTC ACA AAG GAG AAA GTA AAG         1116
Met Pro Leu Ser Val Leu Ile Thr Thr Phe Thr Lys Glu Lys Val Lys
    335                 340                 345

AGA CCT CTT GAA GGC TTT GGG GTA CTC ATT CCA TCT AAG GAG CAA AAG         1164
Arg Pro Leu Glu Gly Phe Gly Val Leu Ile Pro Ser Lys Glu Gln Lys
350                 355                 360                 365
```

```
CAT GGT TTC AAA ACT CTA GGT ACA CTT TTT TCA TCA ATG ATG TTT CCA      1212
His Gly Phe Lys Thr Leu Gly Thr Leu Phe Ser Ser Met Met Phe Pro
                370                 375                 380

GAT CGT TCC CCT AGT GAC GTT CAT CTA TAT ACA ACT TTT ATT GGT GGG      1260
Asp Arg Ser Pro Ser Asp Val His Leu Tyr Thr Thr Phe Ile Gly Gly
            385                 390                 395

AGT AGG AAC CAG GAA CTA GCC AAA GCT TCC ACT GAC GAA TTA AAA CAA      1308
Ser Arg Asn Gln Glu Leu Ala Lys Ala Ser Thr Asp Glu Leu Lys Gln
        400                 405                 410

GTT GTG ACT TCT GAC CTT CAG CGA CTG TTG GGG GTT GAA GGT GAA CCC      1356
Val Val Thr Ser Asp Leu Gln Arg Leu Leu Gly Val Glu Gly Glu Pro
    415                 420                 425

GTG TCT GTC AAC CAT TAC TAT TGG AGG AAA GCA TTC CCG TTG TAT GAC      1404
Val Ser Val Asn His Tyr Tyr Trp Arg Lys Ala Phe Pro Leu Tyr Asp
430                 435                 440                 445

AGC AGC TAT GAC TCA GTC ATG GAA GCA ATT GAC AAG ATG GAG AAT GAT      1452
Ser Ser Tyr Asp Ser Val Met Glu Ala Ile Asp Lys Met Glu Asn Asp
                450                 455                 460

CTA CCT GGG TTC TTC TAT GCA GGT AAT CAT CGA GGG GGG CTC TCT GTT      1500
Leu Pro Gly Phe Phe Tyr Ala Gly Asn His Arg Gly Gly Leu Ser Val
            465                 470                 475

GGG AAA TCA ATA GCA TCA GGT TGC AAA GCA GCT GAC CTT GTG ATC TCA      1548
Gly Lys Ser Ile Ala Ser Gly Cys Lys Ala Ala Asp Leu Val Ile Ser
        480                 485                 490

TAC CTG GAG TCT TGC TCA AAT GAC AAG AAA CCA AAT GAC AGC TTA TAACATTGTC   1603
Tyr Leu Glu Ser Cys Ser Asn Asp Lys Lys Pro Asn Asp Ser Leu
    495                 500                 505

AAGGTTCGTC CCTTTTTATC ACTTACTTTG TAAACTTGTA AAATGCAACA AGCCGCCGTG     1663

CGATTAGCCA ACAACTCAGC AAAACCCAGA TTCTCATAAG GCTCACTAAT TCCAGAATAA     1723

ACTATTTATG TAAAA                                                      1738
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 508 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Ala Ser Gly Ala Val Ala Asp His Gln Ile Glu Ala Val Ser Gly
 1               5                  10                  15

Lys Arg Val Ala Val Val Gly Ala Gly Val Ser Gly Leu Ala Ala Ala
            20                  25                  30

Tyr Lys Leu Lys Ser Arg Gly Leu Asn Val Thr Val Phe Glu Ala Asp
        35                  40                  45
```

Gly Arg Val Gly Gly Lys Leu Arg Ser Val Met Gln Asn Gly Leu Ile
50 55 60

Trp Asp Glu Gly Ala Asn Thr Met Thr Glu Ala Glu Pro Glu Val Gly
65 70 75 80

Ser Leu Leu Asp Asp Leu Gly Leu Arg Glu Lys Gln Gln Phe Pro Ile
85 90 95

Ser Gln Lys Lys Arg Tyr Ile Val Arg Asn Gly Val Pro Val Met Leu
100 105 110

Pro Thr Asn Pro Ile Glu Leu Val Thr Ser Ser Val Leu Ser Thr Gln
115 120 125

Ser Lys Phe Gln Ile Leu Leu Glu Pro Phe Leu Trp Lys Lys Lys Ser
130 135 140

Ser Lys Val Ser Asp Ala Ser Ala Glu Glu Ser Val Ser Glu Phe Phe
145 150 155 160

Gln Arg His Phe Gly Gln Glu Val Val Asp Tyr Leu Ile Asp Pro Phe
165 170 175

Val Gly Gly Thr Ser Ala Ala Asp Pro Asp Ser Leu Ser Met Lys His
180 185 190

Ser Phe Pro Asp Leu Trp Asn Val Glu Lys Ser Phe Gly Ser Ile Ile
195 200 205

Val Gly Ala Ile Arg Thr Lys Phe Ala Ala Lys Gly Gly Lys Ser Arg
210 215 220

Asp Thr Lys Ser Ser Pro Gly Thr Lys Lys Gly Ser Arg Gly Ser Phe
225 230 235 240

Ser Phe Lys Gly Gly Met Gln Ile Leu Pro Asp Thr Leu Cys Lys Ser
245 250 255

Leu Ser His Asp Glu Ile Asn Leu Asp Ser Lys Val Leu Ser Leu Ser
260 265 270

Tyr Asn Ser Gly Ser Arg Gln Glu Asn Trp Ser Leu Ser Cys Val Ser
275 280 285

His Asn Glu Thr Gln Arg Gln Asn Pro His Tyr Asp Ala Val Ile Met
290 295 300

Thr Ala Pro Leu Cys Asn Val Lys Glu Met Lys Val Met Lys Gly Gly
305 310 315 320

Gln Pro Phe Gln Leu Asn Phe Leu Pro Glu Ile Asn Tyr Met Pro Leu
325 330 335

Ser Val Leu Ile Thr Thr Phe Thr Lys Glu Lys Val Lys Arg Pro Leu
340 345 350

Glu Gly Phe Gly Val Leu Ile Pro Ser Lys Glu Gln Lys His Gly Phe
355 360 365

Lys Thr Leu Gly Thr Leu Phe Ser Ser Met Met Phe Pro Asp Arg Ser
370 375 380

Pro Ser Asp Val His Leu Tyr Thr Thr Phe Ile Gly Gly Ser Arg Asn

```
385                 390                 395                 400

Gln Glu Leu Ala Lys Ala Ser Thr Asp Glu Leu Lys Gln Val Val Thr
                405                 410                 415

Ser Asp Leu Gln Arg Leu Leu Gly Val Glu Gly Glu Pro Val Ser Val
            420                 425                 430

Asn His Tyr Tyr Trp Arg Lys Ala Phe Pro Leu Tyr Asp Ser Ser Tyr
        435                 440                 445

Asp Ser Val Met Glu Ala Ile Asp Lys Met Glu Asn Asp Leu Pro Gly
    450                 455                 460

Phe Phe Tyr Ala Gly Asn His Arg Gly Gly Leu Ser Val Gly Lys Ser
465                 470                 475                 480

Ile Ala Ser Gly Cys Lys Ala Ala Asp Leu Val Ile Ser Tyr Leu Glu
                485                 490                 495

Ser Cys Ser Asn Asp Lys Lys Pro Asn Asp Ser Leu
            500                 505
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1698 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 2..1453
(D) OTHER INFORMATION: /note= "Maize protox-1 cDNA (not full-length); sequence from pMDC-4"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
G AAT TCG GCG GAC TGC GTC GTG GTG GGC GGA GGC ATC AGT GGC CTC          46
  Asn Ser Ala Asp Cys Val Val Val Gly Gly Gly Ile Ser Gly Leu
  1               5                  10                  15

TGC ACC GCG CAG GCG CTG GCC ACG CGG CAC GGC GTC GGG GAC GTG CTT        94
Cys Thr Ala Gln Ala Leu Ala Thr Arg His Gly Val Gly Asp Val Leu
            20                  25                  30

GTC ACG GAG GCC CGC GCC CGC CCC GGC GGC AAC ATT ACC ACC GTC GAG       142
Val Thr Glu Ala Arg Ala Arg Pro Gly Gly Asn Ile Thr Thr Val Glu
        35                  40                  45

CGC CCC GAG GAA GGG TAC CTC TGG GAG GAG GGT CCC AAC AGC TTC CAG       190
Arg Pro Glu Glu Gly Tyr Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln
    50                  55                  60
```

```
CCC TCC GAC CCC GTT CTC ACC ATG GCC GTG GAC AGC GGA CTG AAG GAT      238
Pro Ser Asp Pro Val Leu Thr Met Ala Val Asp Ser Gly Leu Lys Asp
     65                  70                  75

GAC TTG GTT TTT GGG GAC CCA AAC GCG CCG CGT TTC GTG CTG TGG GAG      286
Asp Leu Val Phe Gly Asp Pro Asn Ala Pro Arg Phe Val Leu Trp Glu
 80                  85                  90                  95

GGG AAG CTG AGG CCC GTG CCA TCC AAG CCC GCC GAC CTC CCG TTC TTC      334
Gly Lys Leu Arg Pro Val Pro Ser Lys Pro Ala Asp Leu Pro Phe Phe
                100                 105                 110

GAT CTC ATG AGC ATC CCA GGG AAG CTC AGG GCC GGT CTA GGC GCG CTT      382
Asp Leu Met Ser Ile Pro Gly Lys Leu Arg Ala Gly Leu Gly Ala Leu
            115                 120                 125

GGC ATC CGC CCG CCT CCT CCA GGC CGC GAA GAG TCA GTG GAG GAG TTC      430
Gly Ile Arg Pro Pro Pro Pro Gly Arg Glu Glu Ser Val Glu Glu Phe
        130                 135                 140

GTG CGC CGC AAC CTC GGT GCT GAG GTC TTT GAG CGC CTC ATT GAG CCT      478
Val Arg Arg Asn Leu Gly Ala Glu Val Phe Glu Arg Leu Ile Glu Pro
    145                 150                 155

TTC TGC TCA GGT GTC TAT GCT GGT GAT CCT TCT AAG CTC AGC ATG AAG      526
Phe Cys Ser Gly Val Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys
160                 165                 170                 175

GCT GCA TTT GGG AAG GTT TGG CGG TTG GAA GAA ACT GGA GGT AGT ATT      574
Ala Ala Phe Gly Lys Val Trp Arg Leu Glu Glu Thr Gly Gly Ser Ile
                180                 185                 190

ATT GGT GGA ACC ATC AAG ACA ATT CAG GAG AGG AGC AAG AAT CCA AAA      622
Ile Gly Gly Thr Ile Lys Thr Ile Gln Glu Arg Ser Lys Asn Pro Lys
            195                 200                 205

CCA CCG AGG GAT GCC CGC CTT CCG AAG CCA AAA GGG CAG ACA GTT GCA      670
Pro Pro Arg Asp Ala Arg Leu Pro Lys Pro Lys Gly Gln Thr Val Ala
        210                 215                 220

TCT TTC AGG AAG GGT CTT GCC ATG CTT CCA AAT GCC ATT ACA TCC AGC      718
Ser Phe Arg Lys Gly Leu Ala Met Leu Pro Asn Ala Ile Thr Ser Ser
    225                 230                 235

TTG GGT AGT AAA GTC AAA CTA TCA TGG AAA CTC AGG AGC ATT ACA AAA      766
Leu Gly Ser Lys Val Lys Leu Ser Trp Lys Leu Thr Ser Ile Thr Lys
240                 245                 250                 255

TCA GAT GAC AAG GGA TAT GTT TTG GAG TAT GAA ACG CCA GAA GGC GTT      814
Ser Asp Asp Lys Gly Tyr Val Leu Glu Tyr Glu Thr Pro Glu Gly Val
                260                 265                 270

GTT TCG GTG CAG GCT AAA AGT GTT ATC ATG ACT ATT CCA TCA TAT GTT      862
Val Ser Val Gln Ala Lys Ser Val Ile Met Thr Ile Pro Ser Tyr Val
            275                 280                 285

GCT AGC AAC ATT TTG CGT CCA CTT TCA AGC GAT GCT GCA GAT GCT CTA      910
Ala Ser Asn Ile Leu Arg Pro Leu Ser Ser Asp Ala Ala Asp Ala Leu
        290                 295                 300

TCA AGA TTC TAT TAT CCA CCG GTT GCT GCT GTA ACT GTT TCG TAT CCA      958
Ser Arg Phe Tyr Tyr Pro Pro Val Ala Ala Val Thr Val Ser Tyr Pro.
    305                 310                 315
```

```
AAG GAA GCA ATT AGA AAA GAA TGC TTA ATT GAT GGG GAA CTC CAG GGC      1006
Lys Glu Ala Ile Arg Lys Glu Cys Leu Ile Asp Gly Glu Leu Gln Gly
320                 325                 330                 335

TTT GGC CAG TTG CAT CCA CGT AGT CAA GGA GTT GAC ACA TTA GGA ACA      1054
Phe Gly Gln Leu His Pro Arg Ser Gln Gly Val Glu Thr Leu Gly Thr
                340                 345                 350

ATA TAC AGT TCC TCA CTC TTT CCA AAT CGT GCT CCT GAC GGT AGG GTG      1102
Ile Tyr Ser Ser Ser Leu Phe Pro Asn Arg Ala Pro Asp Gly Arg Val
            355                 360                 365

TTA CTT CTA AAC TAC ATA GGA GGT GCT ACA AAC ACA GGA ATT GTT TCC      1150
Leu Leu Leu Asn Tyr Ile Gly Gly Ala Thr Asn Thr Gly Ile Val Ser
        370                 375                 380

AAG ACT GAA AGT GAG CTG GTC GAA GCA GTT GAC CGT GAC CTC CGA AAA      1198
Lys Thr Glu Ser Glu Leu Val Glu Ala Val Asp Arg Asp Leu Arg Lys
    385                 390                 395

ATG CTT ATA AAT TCT ACA GCA GTG GAC CCT TTA GTC CTT GGT GTT CGA      1246
Met Leu Ile Asn Ser Thr Ala Val Asp Pro Leu Val Leu Gly Val Arg
400                 405                 410                 415

GTT TGG CCA CAA GCC ATA CCT CAG TTC CTG GTA GGA CAT CTT GAT CTT      1294
Val Trp Pro Gln Ala Ile Pro Gln Phe Leu Val Gly His Leu Asp Leu
                420                 425                 430

CTG GAA GCC GCA AAA GCT GCC CTG GAC CGA GGT GGC TAC GAT GGG CTG      1342
Leu Glu Ala Ala Lys Ala Ala Leu Asp Arg Gly Gly Tyr Asp Gly Leu
            435                 440                 445

TTC CTA GGA GGG AAC TAT GTT GCA GGA GTT GCC CTG GGC AGA TGC GTT      1390
Phe Leu Gly Gly Asn Tyr Val Ala Gly Val Ala Leu Gly Arg Cys Val
        450                 455                 460

GAG GGC GCG TAT GAA AGT GCC TCG CAA ATA TCT GAC TTC TTG ACC AAG      1438
Glu Gly Ala Tyr Glu Ser Ala Ser Gln Ile Ser Asp Phe Leu Thr Lys
    465                 470                 475

TAT GCC TAC AAG TGATGAAAGA AGTGGAGCGC TACTTGTTAA TCGTTTATGT          1490
Tyr Ala Tyr Lys
480

TGCATAGATG AGGTGCCTCC GGGGAAAAAA AAGCTTGAAT AGTATTTTTT ATTCTTATTT    1550

TGTAAATTGC ATTTCTGTTC TTTTTTCTAT CAGTAATTAG TTATATTTTA GTTCTGTAGG    1610

AGATTGTTCT GTTCACTGCC CTTCAAAAGA AATTTTATTT TTCATTCTTT TATGAGAGCT    1670

GTGCTACTTA AAAAAAAAAA AAAAAAAA                                       1698
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 483 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Asn Ser Ala Asp Cys Val Val Val Gly Gly Gly Ile Ser Gly Leu Cys
1 5 10 15

Thr Ala Gln Ala Leu Ala Thr Arg His Gly Val Gly Asp Val Leu Val
20 25 30

Thr Glu Ala Arg Ala Arg Pro Gly Gly Asn Ile Thr Thr Val Glu Arg
35 40 45

Pro Glu Glu Gly Tyr Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro
50 55 60

Ser Asp Pro Val Leu Thr Met Ala Val Asp Ser Gly Leu Lys Asp Asp
65 70 75 80

Leu Val Phe Gly Asp Pro Asn Ala Pro Arg Phe Val Leu Trp Glu Gly
85 90 95

Lys Leu Arg Pro Val Pro Ser Lys Pro Ala Asp Leu Pro Phe Phe Asp
100 105 110

Leu Met Ser Ile Pro Gly Lys Leu Arg Ala Gly Leu Gly Ala Leu Gly
115 120 125

Ile Arg Pro Pro Pro Pro Gly Arg Glu Glu Ser Val Glu Glu Phe Val
130 135 140

Arg Arg Asn Leu Gly Ala Glu Val Phe Glu Arg Leu Ile Glu Pro Phe
145 150 155 160

Cys Ser Gly Val Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala
165 170 175

Ala Phe Gly Lys Val Trp Arg Leu Glu Glu Thr Gly Gly Ser Ile Ile
180 185 190

Gly Gly Thr Ile Lys Thr Ile Gln Glu Arg Ser Lys Asn Pro Lys Pro
195 200 205

Pro Arg Asp Ala Arg Leu Pro Lys Pro Lys Gly Gln Thr Val Ala Ser
210 215 220

Phe Arg Lys Gly Leu Ala Met Leu Pro Asn Ala Ile Thr Ser Ser Leu
225 230 235 240

Gly Ser Lys Val Lys Leu Ser Trp Lys Leu Thr Ser Ile Thr Lys Ser
245 250 255

Asp Asp Lys Gly Tyr Val Leu Glu Tyr Glu Thr Pro Glu Gly Val Val
260 265 270

Ser Val Gln Ala Lys Ser Val Ile Met Thr Ile Pro Ser Tyr Val Ala
275 280 285

Ser Asn Ile Leu Arg Pro Leu Ser Ser Asp Ala Ala Asp Ala Leu Ser
290 295 300

Arg Phe Tyr Tyr Pro Pro Val Ala Ala Val Thr Val Ser Tyr Pro Lys
305 310 315 320

Glu Ala Ile Arg Lys Glu Cys Leu Ile Asp Gly Glu Leu Gln Gly Phe
325 330 335

Gly Gln Leu His Pro Arg Ser Gln Gly Val Glu Thr Leu Gly Thr Ile

```
            340                 345                 350

Tyr Ser Ser Ser Leu Phe Pro Asn Arg Ala Pro Asp Gly Arg Val Leu
        355                 360                 365

Leu Leu Asn Tyr Ile Gly Gly Ala Thr Asn Thr Gly Ile Val Ser Lys
    370                 375                 380

Thr Glu Ser Glu Leu Val Glu Ala Val Asp Arg Asp Leu Arg Lys Met
385                 390                 395                 400

Leu Ile Asn Ser Thr Ala Val Asp Pro Leu Val Leu Gly Val Arg Val
                405                 410                 415

Trp Pro Gln Ala Ile Pro Gln Phe Leu Val Gly His Leu Asp Leu Leu
            420                 425                 430

Glu Ala Ala Lys Ala Ala Leu Asp Arg Gly Gly Tyr Asp Gly Leu Phe
        435                 440                 445

Leu Gly Gly Asn Tyr Val Ala Gly Val Ala Leu Gly Arg Cys Val Glu
    450                 455                 460

Gly Ala Tyr Glu Ser Ala Ser Gln Ile Ser Asp Phe Leu Thr Lys Tyr
465                 470                 475                 480

Ala Tyr Lys
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2061 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 64..1698
(D) OTHER INFORMATION: /note= "Maize protox-2 cDNA; sequence from pWDC-3"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
CTCTCCTACC TCCACCTCCA CGACAACAAG CAAATCCCCA TCCAGTTCCA AACCCTAACT      60

CAA ATG CTC GCT TTG ACT GCC TCA GCC TCA TCC GCT TCG TCC CAT CCT      108
    Met Leu Ala Leu Thr Ala Ser Ala Ser Ser Ala Ser Ser His Pro
      1               5                  10                  15

TAT CGC CAC GCC TCC GCG CAC ACT CGT CGC CCC CGC CTA CGT GCG GTC      156
Tyr Arg His Ala Ser Ala His Thr Arg Arg Pro Arg Leu Arg Ala Val
                20                  25                  30
```

```
CTC GCG ATG GCG GGC TCC GAC GAC CCC CGT GCA GCG CCC GCC AGA TCG        204
Leu Ala Met Ala Gly Ser Asp Asp Pro Arg Ala Ala Pro Ala Arg Ser
                35                  40                  45

GTC GCC GTC GTC GGC GCC GGG GTC AGC GGG CTC GCG GCG GCG TAC AGG        252
Val Ala Val Val Gly Ala Gly Val Ser Gly Leu Ala Ala Ala Tyr Arg
            50                  55                  60

CTC AGA CAG AGC GGC GTG AAC GTA ACG GTG TTC GAA GCG GCC GAC AGG        300
Leu Arg Gln Ser Gly Val Asn Val Thr Val Phe Glu Ala Ala Asp Arg
        65                  70                  75

GCG GGA GGA AAG ATA CGG ACT AAT TCC GAG GGC GGG TTT GTC TGG GAT        348
Ala Gly Gly Lys Ile Arg Thr Asn Ser Glu Gly Gly Phe Val Trp Asp
    80                  85                  90                  95

GAA GGA GCT AAC ACC ATG ACA GAA GGT GAA TGG GAG GCC AGT AGA CTG        396
Glu Gly Ala Asn Thr Met Thr Glu Gly Glu Trp Glu Ala Ser Arg Leu
                100                 105                 110

ATT GAT GAT CTT GGT CTA CAA GAC AAA CAG CAG TAT CCT AAC TCC CAA        444
Ile Asp Asp Leu Gly Leu Gln Asp Lys Gln Gln Tyr Pro Asn Ser Gln
            115                 120                 125

CAC AAG CGT TAC ATT GTC AAA GAT GGA GCA CCA GCA CTG ATT CCT TCG        492
His Lys Arg Tyr Ile Val Lys Asp Gly Ala Pro Ala Leu Ile Pro Ser
        130                 135                 140

GAT CCC ATT TCG CTA ATG AAA AGC AGT GTT CTT TCG ACA AAA TCA AAG        540
Asp Pro Ile Ser Leu Met Lys Ser Ser Val Leu Ser Thr Lys Ser Lys
    145                 150                 155

ATT GCG TTA TTT TTT GAA CCA TTT CTC TAC AAG AAA GCT AAC ACA AGA        588
Ile Ala Leu Phe Phe Glu Pro Phe Leu Tyr Lys Lys Ala Asn Thr Arg
160                 165                 170                 175

AAC TCT GGA AAA GTG TCT GAG GAG CAC TTG AGT GAG AGT GTT GGG AGC        636
Asn Ser Gly Lys Val Ser Glu Glu His Leu Ser Glu Ser Val Gly Ser
                180                 185                 190

TTC TGT GAA CGC CAC TTT GGA AGA GAA GTT GTT GAC TAT TTT GTT GAT        684
Phe Cys Glu Arg His Phe Gly Arg Glu Val Val Asp Tyr Phe Val Asp
            195                 200                 205

CCA TTT GTA GCT GGA ACA AGT GCA GGA GAT CCA GAG TCA CTA TCT ATT        732
Pro Phe Val Ala Gly Thr Ser Ala Gly Asp Pro Glu Ser Leu Ser Ile
        210                 215                 220

CGT CAT GCA TTC CCA GCA TTG TGG AAT TTG GAA AGA AAG TAT GGT TCA        780
Arg His Ala Phe Pro Ala Leu Trp Asn Leu Glu Arg Lys Tyr Gly Ser
    225                 230                 235

GTT ATT GTT GGT GCC ATC TTG TCT AAG CTA GCA GCT AAA GGT GAT CCA        828
Val Ile Val Gly Ala Ile Leu Ser Lys Leu Ala Ala Lys Gly Asp Pro
240                 245                 250                 255

GTA AAG ACA AGA CAT GAT TCA TCA GGG AAA AGA AGG AAT AGA CGA GTG        876
Val Lys Thr Arg His Asp Ser Ser Gly Lys Arg Arg Asn Arg Arg Val
                260                 265                 270

TCG TTT TCA TTT CAT GGT GGA ATG CAG TCA CTA ATA AAT GCA CTT CAC        924
Ser Phe Ser Phe His Gly Gly Met Gln Ser Leu Ile Asn Ala Leu His
            275                 280                 285
```

```
AAT GAA GTT GGA GAT GAT AAT GTG AAG CTT GGT ACA GAA GTG TTG TCA      972
Asn Glu Val Gly Asp Asp Asn Val Lys Leu Gly Thr Glu Val Leu Ser
        290                 295                 300

TTG GCA TGT ACA TTT GAT GGA GTT CCT GCA CTA GGC AGG TGG TCA ATT     1020
Leu Ala Cys Thr Phe Asp Gly Val Pro Ala Leu Gly Arg Trp Ser Ile
    305                 310                 315

TCT GTT GAT TCG AAG GAT AGC GGT GAC AAG GAC CTT GCT AGT AAC CAA     1068
Ser Val Asp Ser Lys Asp Ser Gly Asp Lys Asp Leu Ala Ser Asn Gln
320                 325                 330                 335

ACC TTT GAT GCT GTT ATA ATG ACA GCT CCA TTG TCA AAT GTC CGG AGG     1116
Thr Phe Asp Ala Val Ile Met Thr Ala Pro Leu Ser Asn Val Arg Arg
                340                 345                 350

ATG AAG TTC ACC AAA GGT GGA GCT CCG GTT GTT CTT GAC TTT CTT CCT     1164
Met Lys Phe Thr Lys Gly Gly Ala Pro Val Val Leu Asp Phe Leu Pro
            355                 360                 365

AAG ATG GAT TAT CTA CCA CTA TCT CTC ATG GTG ACT GCT TTT AAG AAG     1212
Lys Met Asp Tyr Leu Pro Leu Ser Leu Met Val Thr Ala Phe Lys Lys
        370                 375                 380

GAT GAT GTC AAG AAA CCT CTG GAA GGA TTT GGG GTC TTA ATA CCT TAC     1260
Asp Asp Val Lys Lys Pro Leu Glu Gly Phe Gly Val Leu Ile Pro Tyr
    385                 390                 395

AAG GAA CAG CAA AAA CAT GGT CTG AAA ACC CTT GGG ACT CTC TTT TCC     1308
Lys Glu Gln Gln Lys His Gly Leu Lys Thr Leu Gly Thr Leu Phe Ser
400                 405                 410                 415

TCA ATG ATG TTC CCA GAT CGA GCT CCT GAT GAC CAA TAT TTA TAT ACA     1356
Ser Met Met Phe Pro Asp Arg Ala Pro Asp Asp Gln Tyr Leu Tyr Thr
                420                 425                 430

ACA TTT GTT GGG GGT AGC CAC AAT AGA GAT CTT GCT GGA GCT CCA ACG     1404
Thr Phe Val Gly Gly Ser His Asn Arg Asp Leu Ala Gly Ala Pro Thr
            435                 440                 445

TCT ATT CTG AAA CAA CTT GTG ACC TCT GAC CTT AAA AAA CTC TTG GGC     1452
Ser Ile Leu Lys Gln Leu Val Thr Ser Asp Leu Lys Lys Leu Leu Gly
        450                 455                 460

GTA GAG GGG CAA CCA ACT TTT GTC AAG CAT GTA TAC TGG GGA AAT GCT     1500
Val Glu Gly Gln Pro Thr Phe Val Lys His Val Tyr Trp Gly Asn Ala
    465                 470                 475

TTT CCT TTG TAT GGC CAT GAT TAT AGT TCT GTA TTG GAA GCT ATA GAA     1548
Phe Pro Leu Tyr Gly His Asp Tyr Ser Ser Val Leu Glu Ala Ile Glu
480                 485                 490                 495

AAG ATG GAG AAA AAC CTT CCA GGG TTC TTC TAC GCA GGA AAT AGC AAG     1596
Lys Met Glu Lys Asn Leu Pro Gly Phe Phe Tyr Ala Gly Asn Ser Lys
                500                 505                 510

GAT GGG CTT GCT GTT GGA AGT GTT ATA GCT TCA GGA AGC AAG GCT GCT     1644
Asp Gly Leu Ala Val Gly Ser Val Ile Ala Ser Gly Ser Lys Ala Ala
            515                 520                 525

GAC CTT GCA ATC TCA TAT CTT GAA TCT CAC ACC AAG CAT AAT AAT TCA     1692
Asp Leu Ala Ile Ser Tyr Leu Glu Ser His Thr Lys His Asn Asn Ser
        530                 535                 540
```

CAT TGAAAGTGTC TGACCTATCC TCTAGCAGTT GTCGACAAAT TTCTCCAGTT 1745
His
545

CATGTACAGT AGAAAGCTAT GCGTTGCAGT TTCAGAACAT CTTCACTTCT TCAGATATTA 1805

ACCCTTCGTT GAACATCCAC CAGAAAGGTA GTCACATGTG TAAGTGGGAA AATGAGGTTA 1865

AAAACTATTA TGGCGGCCGA AATGTTCCTT TTTGTTTTCC TCACAAGTGG CCTACGACAC 1925

TTGATGTTGG AAATACATTT AAATTTGTTG AATTGTTTGA GAACACATGC GTGACGTGTA 1985

ATATTTGCCT ATTGTGATTT TAGCAGTAGT CTTGGCCAGA TTATGCTTTA CGCCTTTAAA 2045

AAAAAAAAAA AAAAAA 2061

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 544 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Leu Ala Leu Thr Ala Ser Ala Ser Ser Ala Ser Ser His Pro Tyr
1 5 10 15

Arg His Ala Ser Ala His Thr Arg Arg Pro Arg Leu Arg Ala Val Leu
20 25 30

Ala Met Ala Gly Ser Asp Asp Pro Arg Ala Ala Pro Ala Arg Ser Val
35 40 45

Ala Val Val Gly Ala Gly Val Ser Gly Leu Ala Ala Ala Tyr Arg Leu
50 55 60

Arg Gln Ser Gly Val Asn Val Thr Val Phe Glu Ala Ala Asp Arg Ala
65 70 75 80

Gly Gly Lys Ile Arg Thr Asn Ser Glu Gly Gly Phe Val Trp Asp Glu
85 90 95

Gly Ala Asn Thr Met Thr Glu Gly Glu Trp Glu Ala Ser Arg Leu Ile
100 105 110

Asp Asp Leu Gly Leu Gln Asp Lys Gln Gln Tyr Pro Asn Ser Gln His
115 120 125

Lys Arg Tyr Ile Val Lys Asp Gly Ala Pro Ala Leu Ile Pro Ser Asp
130 135 140

Pro Ile Ser Leu Met Lys Ser Ser Val Leu Ser Thr Lys Ser Lys Ile
145 150 155 160

Ala Leu Phe Phe Glu Pro Phe Leu Tyr Lys Lys Ala Asn Thr Arg Asn
165 170 175

Ser Gly Lys Val Ser Glu Glu His Leu Ser Glu Ser Val Gly Ser Phe
180 185 190

Cys Glu Arg His Phe Gly Arg Glu Val Val Asp Tyr Phe Val Asp Pro
195 200 205

Phe Val Ala Gly Thr Ser Ala Gly Asp Pro Glu Ser Leu Ser Ile Arg
210 215 220

His Ala Phe Pro Ala Leu Trp Asn Leu Glu Arg Lys Tyr Gly Ser Val
225 230 235 240

Ile Val Gly Ala Ile Leu Ser Lys Leu Ala Ala Lys Gly Asp Pro Val
245 250 255

Lys Thr Arg His Asp Ser Ser Gly Lys Arg Arg Asn Arg Arg Val Ser
260 265 270

Phe Ser Phe His Gly Gly Met Gln Ser Leu Ile Asn Ala Leu His Asn
275 280 285

Glu Val Gly Asp Asp Asn Val Lys Leu Gly Thr Glu Val Leu Ser Leu
290 295 300

Ala Cys Thr Phe Asp Gly Val Pro Ala Leu Gly Arg Trp Ser Ile Ser
305 310 315 320

Val Asp Ser Lys Asp Ser Gly Asp Lys Asp Leu Ala Ser Asn Gln Thr
325 330 335

Phe Asp Ala Val Ile Met Thr Ala Pro Leu Ser Asn Val Arg Arg Met
340 345 350

Lys Phe Thr Lys Gly Gly Ala Pro Val Val Leu Asp Phe Leu Pro Lys
355 360 365

Met Asp Tyr Leu Pro Leu Ser Leu Met Val Thr Ala Phe Lys Lys Asp
370 375 380

Asp Val Lys Lys Pro Leu Glu Gly Phe Gly Val Leu Ile Pro Tyr Lys
385 390 395 400

Glu Gln Gln Lys His Gly Leu Lys Thr Leu Gly Thr Leu Phe Ser Ser
405 410 415

Met Met Phe Pro Asp Arg Ala Pro Asp Asp Gln Tyr Leu Tyr Thr Thr
420 425 430

Phe Val Gly Gly Ser His Asn Arg Asp Leu Ala Gly Ala Pro Thr Ser
435 440 445

Ile Leu Lys Gln Leu Val Thr Ser Asp Leu Lys Lys Leu Leu Gly Val
450 455 460

Glu Gly Gln Pro Thr Phe Val Lys His Val Tyr Trp Gly Asn Ala Phe
465 470 475 480

Pro Leu Tyr Gly His Asp Tyr Ser Ser Val Leu Glu Ala Ile Glu Lys
485 490 495

Met Glu Lys Asn Leu Pro Gly Phe Phe Tyr Ala Gly Asn Ser Lys Asp
500 505 510

Gly Leu Ala Val Gly Ser Val Ile Ala Ser Gly Ser Lys Ala Ala Asp
515 520 525

Leu Ala Ile Ser Tyr Leu Glu Ser His Thr Lys His Asn Asn Ser His

530 535 540

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1811 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 3..1589
(D) OTHER INFORMATION: /product= "wheat protox-1 cDNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
GC GCA ACA ATG GCC ACC GCC ACC GTC GCG GCC GCG TCG CCG CTC CGC          47
   Ala Thr Met Ala Thr Ala Thr Val Ala Ala Ala Ser Pro Leu Arg
     1               5                  10                  15

GGC AGG GTC ACC GGG CGC CCA CAC CGC GTC CGC CCG CGT TGC GCT ACC          95
Gly Arg Val Thr Gly Arg Pro His Arg Val Arg Pro Arg Cys Ala Thr
                20                  25                  30

GCG AGC AGC GCG ACC GAG ACT CCG GCG GCG CCC GGC GTG CGG CTG TCC         143
Ala Ser Ser Ala Thr Glu Thr Pro Ala Ala Pro Gly Val Arg Leu Ser
            35                  40                  45

GCG GAA TGC GTC ATT GTG GGC GCC GGC ATC AGC GGC CTC TGC ACC GCG         191
Ala Glu Cys Val Ile Val Gly Ala Gly Ile Ser Gly Leu Cys Thr Ala
        50                  55                  60

CAG GCG CTG GCC ACC CGA TAC GGC GTC AGC GAC CTG CTC GTC ACG GAG         239
Gln Ala Leu Ala Thr Arg Tyr Gly Val Ser Asp Leu Leu Val Thr Glu
    65                  70                  75

GCC CGC GAC CGC CCG GGC GGC AAC ATC ACC ACC GTC GAG CGT CCC GAC         287
Ala Arg Asp Arg Pro Gly Gly Asn Ile Thr Thr Val Glu Arg Pro Asp
80                  85                  90                  95

GAG GGG TAC CTG TGG GAG GAG GGA CCC AAC AGC TTC CAG CCC TCC GAC         335
Glu Gly Tyr Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp
                100                 105                 110

CCG GTC CTC ACC ATG GCC GTG GAC AGC GGG CTC AAG GAT GAC TTG GTG         383
Pro Val Leu Thr Met Ala Val Asp Ser Gly Leu Lys Asp Asp Leu Val
            115                 120                 125

TTC GGG GAC CCC AAC GCG CCC CGG TTC GTG CTG TGG GAG GGG AAG CTG         431
Phe Gly Asp Pro Asn Ala Pro Arg Phe Val Leu Trp Glu Gly Lys Leu
        130                 135                 140

AGG CCG GTG CCG TCG AAG CCA GGC GAC CTG CCT TTC TTC AGC CTC ATG         479
Arg Pro Val Pro Ser Lys Pro Gly Asp Leu Pro Phe Phe Ser Leu Met
    145                 150                 155
```

```
AGT ATC CCT GGG AAG CTC AGG GCC GGC CTT GGC GCG CTC GGC ATT CGC          527
Ser Ile Pro Gly Lys Leu Arg Ala Gly Leu Gly Ala Leu Gly Ile Arg
160                 165                 170                 175

CCA CCT CCT CCA GGG CGC GAG GAG TCG GTG GAG GAG TTT GTG CGC CGC          575
Pro Pro Pro Pro Gly Arg Glu Glu Ser Val Glu Glu Phe Val Arg Arg
                180                 185                 190

AAC CTC GGT GCC GAG GTC TTT GAG CGC CTC ATC GAG CCT TTC TGC TCA          623
Asn Leu Gly Ala Glu Val Phe Glu Arg Leu Ile Glu Pro Phe Cys Ser
            195                 200                 205

GGT GTA TAT GCT GGT GAT CCT TCG AAG CTT AGT ATG AAG GCT GCA TTT          671
Gly Val Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala Ala Phe
        210                 215                 220

GGG AAG GTC TGG AGG TTG GAG GAG ATT GGA GGT AGT ATT ATT GGT GGA          719
Gly Lys Val Trp Arg Leu Glu Glu Ile Gly Gly Ser Ile Ile Gly Gly
    225                 230                 235

ACC ATC AAG GCG ATT CAG GAT AAA GGG AAG AAC CCC AAA CCG CCA AGG          767
Thr Ile Lys Ala Ile Gln Asp Lys Gly Lys Asn Pro Lys Pro Pro Arg
240                 245                 250                 255

GAT CCC CGA CTT CCG GCA CCA AAG GGA CAG ACG GTG GCA TCT TTC AGG          815
Asp Pro Arg Leu Pro Ala Pro Lys Gly Gln Thr Val Ala Ser Phe Arg
                260                 265                 270

AAG GGT CTA GCC ATG CTC CCG AAT GCC ATC GCA TCT AGG CTG GGT AGT          863
Lys Gly Leu Ala Met Leu Pro Asn Ala Ile Ala Ser Arg Leu Gly Ser
            275                 280                 285

AAA GTC AAG CTG TCA TGG AAG CTT ACG AGC ATT ACA AAG GCG GAC AAC          911
Lys Val Lys Leu Ser Trp Lys Leu Thr Ser Ile Thr Lys Ala Asp Asn
        290                 295                 300

CAA GGA TAT GTA TTA GGT TAT GAA ACA CCA GAA GGA CTT GTT TCA GTG          959
Gln Gly Tyr Val Leu Gly Tyr Glu Thr Pro Glu Gly Leu Val Ser Val
    305                 310                 315

CAG GCT AAA AGT GTT ATC ATG ACC ATC CCG TCA TAT GTT GCT AGT GAT         1007
Gln Ala Lys Ser Val Ile Met Thr Ile Pro Ser Tyr Val Ala Ser Asp
320                 325                 330                 335

ATC TTG CGC CCA CTT TCA ATT GAT GCA GCA GAT GCA CTC TCA AAA TTC         1055
Ile Leu Arg Pro Leu Ser Ile Asp Ala Ala Asp Ala Leu Ser Lys Phe
                340                 345                 350

TAT TAT CCG CCA GTT GCT GCT GTA ACT GTT TCA TAT CCA AAA GAA GCT         1103
Tyr Tyr Pro Pro Val Ala Ala Val Thr Val Ser Tyr Pro Lys Glu Ala
            355                 360                 365

ATT AGA AAA GAA TGC TTA ATT GAT GGG GAG CTC CAG GGT TTC GGC CAG         1151
Ile Arg Lys Glu Cys Leu Ile Asp Gly Glu Leu Gln Gly Phe Gly Gln
        370                 375                 380

TTG CAT CCA CGT AGC CAA GGA GTC GAG ACT TTA GGG ACA ATA TAT AGC         1199
Leu His Pro Arg Ser Gln Gly Val Glu Thr Leu Gly Thr Ile Tyr Ser
    385                 390                 395

TCT TCT CTC TTT CCT AAT CGT GCT CCT GCT GGA AGA GTG TTA CTT CTG         1247
Ser Ser Leu Phe Pro Asn Arg Ala Pro Ala Gly Arg Val Leu Leu Leu
400                 405                 410                 415
```

```
AAC TAT ATC GGG GGT TCT ACA AAT ACA GGG ATC GTC TCC AAG ACT GAG      1295
Asn Tyr Ile Gly Gly Ser Thr Asn Thr Gly Ile Val Ser Lys Thr Glu
                420                 425                 430

AGT GAC TTA GTA GGA GCC GTT GAC CGT GAC CTC AGA AAA ATG TTG ATA      1343
Ser Asp Leu Val Gly Ala Val Asp Arg Asp Leu Arg Lys Met Leu Ile
            435                 440                 445

AAC CCT AGA GCA GCA GAC CCT TTA GCA TTA GGG GTT CGA GTG TGG CCA      1391
Asn Pro Arg Ala Ala Asp Pro Leu Ala Leu Gly Val Arg Val Trp Pro
        450                 455                 460

CAA GCA ATA CCA CAG TTT TTG ATT GGG CAC CTT GAT CGC CTT GCT GCT      1439
Gln Ala Ile Pro Gln Phe Leu Ile Gly His Leu Asp Arg Leu Ala Ala
    465                 470                 475

GCA AAA TCT GCA CTG GGC CAA GGC GGC TAC GAC GGG TTG TTC CTA GGA      1487
Ala Lys Ser Ala Leu Gly Gln Gly Gly Tyr Asp Gly Leu Phe Leu Gly
480                 485                 490                 495

GGA AAC TAC GTC GCA GGA GTT GCC TTG GGC CGA TGC ATC GAG GGT GCG      1535
Gly Asn Tyr Val Ala Gly Val Ala Leu Gly Arg Cys Ile Glu Gly Ala
                500                 505                 510

TAC GAG AGT GCC TCA CAA GTA TCT GAC TTC TTG ACC AAG TAT GCC TAC      1583
Tyr Glu Ser Ala Ser Gln Val Ser Asp Phe Leu Thr Lys Tyr Ala Tyr
            515                 520                 525

AAG TGA TGGAAGTAGT GCATCTCTTC ATTTTGTTGC ATATACGAGG TGAGGCTAGG      1639
Lys

ATCGGTAAAA CATCATGAGA TTCTGTAGTG TTTCTTTAAT TGAAAAAACA AATTTTAGTG   1699

ATGCAATATG TGCTCTTTCC TGTAGTTCGA GCATGTACAT CGGTATGGGA TAAAGTAGAA   1759

TAAGCTATTC TGCAAAAGCA GTGATTTTTT TTGAAAAAAA AAAAAAAAAA AA            1811
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 528 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Ala Thr Met Ala Thr Ala Thr Val Ala Ala Ala Ser Pro Leu Arg Gly
  1               5                  10                  15

Arg Val Thr Gly Arg Pro His Arg Val Arg Pro Arg Cys Ala Thr Ala
            20                  25                  30

Ser Ser Ala Thr Glu Thr Pro Ala Ala Pro Gly Val Arg Leu Ser Ala
        35                  40                  45

Glu Cys Val Ile Val Gly Ala Gly Ile Ser Gly Leu Cys Thr Ala Gln
    50                  55                  60

Ala Leu Ala Thr Arg Tyr Gly Val Ser Asp Leu Leu Val Thr Glu Ala
```

65 70 75 80

Arg Asp Arg Pro Gly Gly Asn Ile Thr Thr Val Glu Arg Pro Asp Glu
85 90 95

Gly Tyr Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp Pro
100 105 110

Val Leu Thr Met Ala Val Asp Ser Gly Leu Lys Asp Asp Leu Val Phe
115 120 125

Gly Asp Pro Asn Ala Pro Arg Phe Val Leu Trp Glu Gly Lys Leu Arg
130 135 140

Pro Val Pro Ser Lys Pro Gly Asp Leu Pro Phe Phe Ser Leu Met Ser
145 150 155 160

Ile Pro Gly Lys Leu Arg Ala Gly Leu Gly Ala Leu Gly Ile Arg Pro
165 170 175

Pro Pro Pro Gly Arg Glu Glu Ser Val Glu Glu Phe Val Arg Arg Asn
180 185 190

Leu Gly Ala Glu Val Phe Glu Arg Leu Ile Glu Pro Phe Cys Ser Gly
195 200 205

Val Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala Ala Phe Gly
210 215 22

Lys Val Trp Arg Leu Glu Glu Ile Gly Gly Ser Ile Ile Gly Gly Thr
225 230 235 240

Ile Lys Ala Ile Gln Asp Lys Gly Lys Asn Pro Lys Pro Pro Arg Asp
245 250 255

Pro Arg Leu Pro Ala Pro Lys Gly Gln Thr Val Ala Ser Phe Arg Lys
260 265 270

Gly Leu Ala Met Leu Pro Asn Ala Ile Ala Ser Arg Leu Gly Ser Lys
275 280 285

Val Lys Leu Ser Trp Lys Leu Thr Ser Ile Thr Lys Ala Asp Asn Gln
290 295 300

Gly Tyr Val Leu Gly Tyr Glu Thr Pro Glu Gly Leu Val Ser Val Gln
305 310 315 320

Ala Lys Ser Val Ile Met Thr Ile Pro Ser Tyr Val Ala Ser Asp Ile
325 330 335

Leu Arg Pro Leu Ser Ile Asp Ala Ala Asp Ala Leu Ser Lys Phe Tyr
340 345 350

Tyr Pro Pro Val Ala Ala Val Thr Val Ser Tyr Pro Lys Glu Ala Ile
355 360 365

Arg Lys Glu Cys Leu Ile Asp Gly Glu Leu Gln Gly Phe Gly Gln Leu
370 375 380

His Pro Arg Ser Gln Gly Val Glu Thr Leu Gly Thr Ile Tyr Ser Ser
385 390 395 400

Ser Leu Phe Pro Asn Arg Ala Pro Ala Gly Arg Val Leu Leu Leu Asn
405 410 415

```
Tyr Ile Gly Gly Ser Thr Asn Thr Gly Ile Val Ser Lys Thr Glu Ser
            420                 425                 430

Asp Leu Val Gly Ala Val Asp Arg Asp Leu Arg Lys Met Leu Ile Asn
        435                 440                 445

Pro Arg Ala Ala Asp Pro Leu Ala Leu Gly Val Arg Val Trp Pro Gln
    450                 455                 460

Ala Ile Pro Gln Phe Leu Ile Gly His Leu Asp Arg Leu Ala Ala Ala
465                 470                 475                 480

Lys Ser Ala Leu Gly Gln Gly Gly Tyr Asp Gly Leu Phe Leu Gly Gly
                485                 490                 495

Asn Tyr Val Ala Gly Val Ala Leu Gly Arg Cys Ile Glu Gly Ala Tyr
            500                 505                 510

Glu Ser Ala Ser Gln Val Ser Asp Phe Leu Thr Lys Tyr Ala Tyr Lys
        515                 520                 525
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1847 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 55..1683
(D) OTHER INFORMATION: /product= "soybean protox-1 cDNA"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
CTTTAGCACA GTGTTGAAGA TAACGAACGA ATAGTGCCAT TACTGTAACC AACC ATG          57
                                                          Met
                                                           1

GTT TCC GTC TTC AAC GAG ATC CTA TTC CCG CCG AAC CAA ACC CTT CTT          105
Val Ser Val Phe Asn Glu Ile Leu Phe Pro Pro Asn Gln Thr Leu Leu
          5                  10                  15

CGC CCC TCC CTC CAT TCC CCA ACC TCT TTC TTC ACC TCT CCC ACT CGA          153
Arg Pro Ser Leu His Ser Pro Thr Ser Phe Phe Thr Ser Pro Thr Arg
      20                  25                  30

AAA TTC CCT CGC TCT CGC CCT AAC CCT ATT CTA CGC TGC TCC ATT GCG          201
Lys Phe Pro Arg Ser Arg Pro Asn Pro Ile Leu Arg Cys Ser Ile Ala
  35                  40                  45

GAG GAA TCC ACC GCG TCT CCG CCC AAA ACC AGA GAC TCC GCC CCC GTG          249
Glu Glu Ser Thr Ala Ser Pro Pro Lys Thr Arg Asp Ser Ala Pro Val
50                  55                  60                  65
```

```
GAC TGC GTC GTC GTC GGC GGA GGC GTC AGC GGC CTC TGC ATC GCC CAG        297
  p Cys Val Val Val Gly Gly Gly Val Ser Gly Leu Cys Ile Ala Gln
                 70                  75                  80

GCC CTC GCC ACC AAA CAC GCC AAT GCC AAC GTC GTC GTC ACG GAG GCC        345
Ala Leu Ala Thr Lys His Ala Asn Ala Asn Val Val Val Thr Glu Ala
            85                  90                  95

CGA GAC CGC GTC GGC GGC AAC ATC ACC ACG ATG GAG AGG GAC GGA TAC        393
Arg Asp Arg Val Gly Gly Asn Ile Thr Thr Met Glu Arg Asp Gly Tyr
        100                 105                 110

CTC TGG GAA GAA GGC CCC AAC AGC TTC CAG CCT TCT GAT CCA ATG CTC        441
Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp Pro Met Leu
    115                 120                 125

ACC ATG GTG GTG GAC AGT GGT TTA AAG GAT GAG CTT GTT TTG GGG GAT        489
Thr Met Val Val Asp Ser Gly Leu Lys Asp Glu Leu Val Leu Gly Asp
130                 135                 140                 145

CCT GAT GCA CCT CGG TTT GTG TTG TGG AAC AGG AAG TTG AGG CCG GTG        537
Pro Asp Ala Pro Arg Phe Val Leu Trp Asn Arg Lys Leu Arg Pro Val
                150                 155                 160

CCC GGG AAG CTG ACT GAT TTG CCT TTC TTT GAC TTG ATG AGC ATT GGT        585
Pro Gly Lys Leu Thr Asp Leu Pro Phe Phe Asp Leu Met Ser Ile Gly
            165                 170                 175

GGC AAA ATC AGG GCT GGC TTT GGT GCG CTT GGA ATT CGG CCT CCT CCT        633
Gly Lys Ile Arg Ala Gly Phe Gly Ala Leu Gly Ile Arg Pro Pro Pro
        180                 185                 190

CCA GGT CAT GAG GAA TCG GTT GAA GAG TTT GTT CGT CGG AAC CTT GGT        681
Pro Gly His Glu Glu Ser Val Glu Glu Phe Val Arg Arg Asn Leu Gly
    195                 200                 205

GAT GAG GTT TTT GAA CGG TTG ATA GAG CCT TTT TGT TCA GGG GTC TAT        729
Asp Glu Val Phe Glu Arg Leu Ile Glu Pro Phe Cys Ser Gly Val Tyr
210                 215                 220                 225

GCA GGC GAT CCT TCA AAA TTA AGT ATG AAA GCA GCA TTC GGG AAA GTT        777
Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala Ala Phe Gly Lys Val
                230                 235                 240

TGG AAG CTG GAA AAA AAT GGT GGT AGC ATT ATT GGT GGA ACT TTC AAA        825
Trp Lys Leu Glu Lys Asn Gly Gly Ser Ile Ile Gly Gly Thr Phe Lys
            245                 250                 255

GCA ATA CAA GAG AGA AAT GGA GCT TCA AAA CCA CCT CGA GAT CCG CGT        873
Ala Ile Gln Glu Arg Asn Gly Ala Ser Lys Pro Pro Arg Asp Pro Arg
        260                 265                 270

CTG CCA AAA CCA AAA GGT CAG ACT GTT GGA TCT TTC CGG AAG GGA CTT        921
Leu Pro Lys Pro Lys Gly Gln Thr Val Gly Ser Phe Arg Lys Gly Leu
    275                 280                 285

ACC ATG TTG CCT GAT GCA ATT TCT GCC AGA CTA GGC AAC AAA GTA AAG        969
Thr Met Leu Pro Asp Ala Ile Ser Ala Arg Leu Gly Asn Lys Val Lys
290                 295                 300                 305

TTA TCT TGG AAG CTT TCA AGT ATT AGT AAA CTG GAT AGT GGA GAG TAC       1017
Leu Ser Trp Lys Leu Ser Ser Ile Ser Lys Leu Asp Ser Gly Glu Tyr
                310                 315                 320
```

```
AGT TTG ACA TAT GAA ACA CCA GAA GGA GTG GTT TCT TTG CAG TGC AAA      1065
Ser Leu Thr Tyr Glu Thr Pro Glu Gly Val Val Ser Leu Gln Cys Lys
            325                 330                 335

ACT GTT GTC CTG ACC ATT CCT TCC TAT GTT GCT AGT ACA TTG CTG CGT      1113
Thr Val Val Leu Thr Ile Pro Ser Tyr Val Ala Ser Thr Leu Leu Arg
        340                 345                 350

CCT CTG TCT GCT GCT GCT GCA GAT GCA CTT TCA AAG TTT TAT TAC CCT      1161
Pro Leu Ser Ala Ala Ala Ala Asp Ala Leu Ser Lys Phe Tyr Tyr Pro
    355                 360                 365

CCA GTT GCT GCA GTT TCC ATA TCC TAT CCA AAA GAA GCT ATT AGA TCA      1209
Pro Val Ala Ala Val Ser Ile Ser Tyr Pro Lys Glu Ala Ile Arg Ser
370                 375                 380                 385

GAA TGC TTG ATA GAT GGT GAG TTG AAG GGG TTT GGT CAA TTG CAT CCA      1257
Glu Cys Leu Ile Asp Gly Glu Leu Lys Gly Phe Gly Gln Leu His Pro
            390                 395                 400

CGT AGC CAA GGA GTG GAA ACA TTA GGA ACT ATA TAC AGC TCA TCA CTA      1305
Arg Ser Gln Gly Val Glu Thr Leu Gly Thr Ile Tyr Ser Ser Ser Leu
        405                 410                 415

TTC CCC AAC CGA GCA CCA CCT GGA AGG GTT CTA CTC TTG AAT TAC ATT      1353
Phe Pro Asn Arg Ala Pro Pro Gly Arg Val Leu Leu Leu Asn Tyr Ile
    420                 425                 430

GGA GGA GCA ACT AAT ACT GGA ATT TTA TCG AAG ACG GAC AGT GAA CTT      1401
Gly Gly Ala Thr Asn Thr Gly Ile Leu Ser Lys Thr Asp Ser Glu Leu
435                 440                 445

GTG GAA ACA GTT GAT CGA GAT TTG AGG AAA ATC CTT ATA AAC CCA AAT      1449
Val Glu Thr Val Asp Arg Asp Leu Arg Lys Ile Leu Ile Asn Pro Asn
450                 455                 460                 465

GCC CAG GAT CCA TTT GTA GTG GGG GTG AGA CTG TGG CCT CAA GCT ATT      1497
Ala Gln Asp Pro Phe Val Val Gly Val Arg Leu Trp Pro Gln Ala Ile
            470                 475                 480

CCA CAG TTC TTA GTT GGC CAT CTT GAT CTT CTA GAT GTT GCT AAA GCT      1545
Pro Gln Phe Leu Val Gly His Leu Asp Leu Leu Asp Val Ala Lys Ala
        485                 490                 495

TCT ATC AGA AAT ACT GGG TTT GAA GGG CTC TTC CTT GGG GGT AAT TAT      1593
Ser Ile Arg Asn Thr Gly Phe Glu Gly Leu Phe Leu Gly Gly Asn Tyr
    500                 505                 510

GTG TCT GGT GTT GCC TTG GGA CGA TGC GTT GAG GGA GCC TAT GAG GTA      1641
Val Ser Gly Val Ala Leu Gly Arg Cys Val Glu Gly Ala Tyr Glu Val
515                 520                 525

GCA GCT GAA GTA AAC GAT TTT CTC ACA AAT AGA GTG TAC AAA              1683
Ala Ala Glu Val Asn Asp Phe Leu Thr Asn Arg Val Tyr Lys
530                 535                 540

TAGTAGCAGT TTTTGTTTTT GTGGTGGAAT GGGTGATGGG ACTCTCGTGT TCCATTGAAT    1743

TATAATAATG TGAAAGTTTC TCAAATTCGT TCGATAGGTT TTTGGCGGCT TCTATTGCTG    1803

ATAATGTAAA ATCCTCTTTA AGTTTGAAAA AAAAAAAAAA AAAA                      1847
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 543 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID 12:

```
Met Val Ser Val Phe Asn Glu Ile Leu Phe Pro Pro Asn Gln Thr Leu
 1               5                  10                  15

Leu Arg Pro Ser Leu His Ser Pro Thr Ser Phe Phe Thr Ser Pro Thr
            20                  25                  30

Arg Lys Phe Pro Arg Ser Arg Pro Asn Pro Ile Leu Arg Cys Ser Ile
        35                  40                  45

Ala Glu Glu Ser Thr Ala Ser Pro Pro Lys Thr Arg Asp Ser Ala Pro
    50                  55                  60

Val Asp Cys Val Val Val Gly Gly Gly Val Ser Gly Leu Cys Ile Ala
65                  70                  75                  80

Gln Ala Leu Ala Thr Lys His Ala Asn Ala Asn Val Val Val Thr Glu
                85                  90                  95

Ala Arg Asp Arg Val Gly Gly Asn Ile Thr Thr Met Glu Arg Asp Gly
            100                 105                 110

Tyr Leu Trp Glu Glu Gly Pro Asn Ser Phe Gln Pro Ser Asp Pro Met
        115                 120                 125

Leu Thr Met Val Val Asp Ser Gly Leu Lys Asp Glu Leu Val Leu Gly
    130                 135                 140

Asp Pro Asp Ala Pro Arg Phe Val Leu Trp Asn Arg Lys Leu Arg Pro
145                 150                 155                 160

Val Pro Gly Lys Leu Thr Asp Leu Pro Phe Phe Asp Leu Met Ser Ile
                165                 170                 175

Gly Gly Lys Ile Arg Ala Gly Phe Gly Ala Leu Gly Ile Arg Pro Pro
            180                 185                 190

Pro Pro Gly His Glu Glu Ser Val Glu Glu Phe Val Arg Arg Asn Leu
        195                 200                 205

Gly Asp Glu Val Phe Glu Arg Leu Ile Glu Pro Phe Cys Ser Gly Val
    210                 215                 220

Tyr Ala Gly Asp Pro Ser Lys Leu Ser Met Lys Ala Ala Phe Gly Lys
225                 230                 235                 240

Val Trp Lys Leu Glu Lys Asn Gly Gly Ser Ile Ile Gly Gly Thr Phe
                245                 250                 255

Lys Ala Ile Gln Glu Arg Asn Gly Ala Ser Lys Pro Pro Arg Asp Pro
            260                 265                 270

Arg Leu Pro Lys Pro Lys Gly Gln Thr Val Gly Ser Phe Arg Lys Gly
        275                 280                 285
```

Leu Thr Met Leu Pro Asp Ala Ile Ser Ala Arg Leu Gly Asn Lys Val
290 295 300

Lys Leu Ser Trp Lys Leu Ser Ser Ile Ser Lys Leu Asp Ser Gly Glu
305 310 315 320

Tyr Ser Leu Thr Tyr Glu Thr Pro Glu Gly Val Val Ser Leu Gln Cys
325 330 335

Lys Thr Val Val Leu Thr Ile Pro Ser Tyr Val Ala Ser Thr Leu Leu
340 345 350

Arg Pro Leu Ser Ala Ala Ala Ala Asp Ala Leu Ser Lys Phe Tyr Tyr
355 360 365

Pro Pro Val Ala Ala Val Ser Ile Ser Tyr Pro Lys Glu Ala Ile Arg
370 375 380

Ser Glu Cys Leu Ile Asp Gly Glu Leu Lys Gly Phe Gly Gln Leu His
385 390 395 400

Pro Arg Ser Gln Gly Val Glu Thr Leu Gly Thr Ile Tyr Ser Ser Ser
405 410 415

Leu Phe Pro Asn Arg Ala Pro Pro Gly Arg Val Leu Leu Leu Asn Tyr
420 425 430

Ile Gly Gly Ala Thr Asn Thr Gly Ile Leu Ser Lys Thr Asp Ser Glu
435 440 445

Leu Val Glu Thr Val Asp Arg Asp Leu Arg Lys Ile Leu Ile Asn Pro
450 455 460

Asn Ala Gln Asp Pro Phe Val Val Gly Val Arg Leu Trp Pro Gln Ala
465 470 475 480

Ile Pro Gln Phe Leu Val Gly His Leu Asp Leu Leu Asp Val Ala Lys
485 490 495

Ala Ser Ile Arg Asn Thr Gly Phe Glu Gly Leu Phe Leu Gly Gly Asn
500 505 510

Tyr Val Ser Gly Val Ala Leu Gly Arg Cys Val Glu Gly Ala Tyr Glu
515 520 525

Val Ala Ala Glu Val Asn Asp Phe Leu Thr Asn Arg Val Tyr Lys
530 535 540

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 583 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(ix) FEATURE:
(A) NAME/KEY: promoter
(B) LOCATION: 1..583

(D) OTHER INFORMATION: /function= "arabidopsis protox-1 promoter"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAATTCCGAT CGAATTATAT AATTATCATA AATTTGAATA AGCATGTTGC CTTTTATTAA 60

AGAGGTTTAA TAAAGTTTGG TAATAATGGA CTTTGACTTC AAACTCGATT CTCATGTAAT 120

TAATTAATAT TTACATCAAA ATTTGGTCAC TAATATTACC AAATTAATAT ACTAAAATGT 180

TAATTCGCAA ATAAAACACT AATTCCAAAT AAAGGGTCAT TATGATAAAC ACGTATTGAA 240

CTTGATAAAG CAAAGCAAAA ATAATGGGTT TCAAGGTTTG GGTTATATAT GACAAAAAAA 300

AAAAAAGGTT TGGTTATATA TCTATTGGGC CTATAACCAT GTTATACAAA TTTGGGCCTA 360

ACTAAAATAA TAAAATAAAC GTAATGGTCC TTTTTATATT TGGGTCAAAC CCAACTCTAA 420

ACCCAAACCA AAGAAAAAGT ATACGGTACG GTACACAGAC TTATGGTGTG TGTGATTGCA 480

GGTGAATATT TCTCGTCGTC TTCTCCTTTC TTCTGAAGAA GATTACCCAA TCTGAAAAAA 540

ACCAAGAAGC TGACAAAATT CCGAATTCTC TGCGATTTCC ATG 583

The invention as described herein is contemplated to include the following enumerated embodiments:

1. An isolated DNA molecule encoding a plant protoporphyrinogen oxidase(protox) enzyme selected from the group consisting of a soybean protox enzyme and a wheat protox enzyme.

2. The isolated DNA molecule of claim 1 encoding said soybean protox enzyme comprising the amino acid sequence set forth in SEQ ID No.12.

3. The isolated DNA molecule of claim 2 comprising the nucleotide sequence set forth in SEQ ID No. 11.

4. The isolated DNA molecule of claim 1, encoding said wheat protox enzyme comprising the amino acid sequence set forth in SEQ ID No. 10.

5. The isolated DNA molecule of claim 4 comprising the nucleotide sequence set forth in SEQ ID No. 9.

6. A DNA molecule encoding a modified protoporphyrinogen oxidase(protox) comprising a plant protox wherein the cysteine occurring at the position corresponding to amino acid 161 of SEQ ID No. 6 is replaced with another amino acid, wherein said modified protox is tolerant to a herbicide in amounts which inhibit said plant protox.

7. The DNA molecule of claim 6 wherein said cysteine is replaced with a phenylalanine.

8. A DNA molecule encoding a modified protoporphyrinogen oxidase(protox) comprising a plant protox wherein the isoleucine occurring at the position corresponding to amino acid 421 of SEQ ID No. 6 is replaced with another amino acid, wherein said modified protox is tolerant to a herbicide in amounts which inhibit said plant protox.

9. The DNA molecule of claim 8 wherein said isoleucine is replaced with a threonine.

10. A DNA molecule encoding a modified protoporphyrinogen oxidase(protox) comprising a plant protox having a first amino acid substitution and a second amino acid substitution,

said first amino acid substitution having the property of conferring resistance to a protox inhibitor; and

said second amino acid substitution having the property of enhancing said resistance conferred by said first amino acid substitution.

11. The DNA molecule of claim 10 wherein said second amino acid substitution occurs at a position selected from the group consisting of

(i) the position corresponding to the serine at amino acid 305 of SEQ ID NO. 2;

(ii) the position corresponding to the threonine at amino acid 249 of SEQ ID NO. 2;

(iii) the position corresponding to the proline at amino acid 118 of SEQ ID NO. 2;

(iv) the position corresponding to the asparagine at amino acid 425 of SEQ ID NO. 2; and

(v) the position corresponding to the tyrosine at amino acid 498 of SEQ ID NO. 2.

12. The DNA molecule of claim 11, wherein said first amino acid substitution occurs at a position selected from the group consisting of

(a) the position corresponding to the alanine at amino acid 166 of SEQ ID No. 6;

(b) the position corresponding to the glycine at position 167 of SEQ ID No. 6;

(c) the position corresponding to the tyrosine at amino acid 372 of SEQ ID No. 6;

(d) the position corresponding to the cysteine at amino acid 161 of SEQ ID No. 6; and

(e) the position corresponding to the isoleucine at amino acid 421 of SEQ ID No. 6.

13. The DNA molecule of claim 11 wherein said second amino acid substitution occurs at the position corresponding to the serine at amino acid 305 of SEQ ID NO. 2 and said first amino acid substitution occurs at a position selected from the group consisting of

(a) the position corresponding to the alanine at amino acid 166 of SEQ ID No. 6; and

(b) the position corresponding to the tyrosine at amino acid 372 of SEQ ID No. 6.

14 The DNA molecule of claim 13 wherein said serine occurring at the position corresponding to amino acid 305 of SEQ ID NO. 2 is replaced with leucine.

15. The DNA molecule of claim 11 wherein said second amino acid substitution occurs at the position corresponding to the threonine at amino acid 249 of SEQ ID NO. 2 and said first amino acid substitution occurs at a position selected from the group consisting of

(a) the position corresponding to the alanine at amino acid 166 of SEQ ID No. 6; and

(b) the position corresponding to the tyrosine at amino acid 372 of SEQ ID No. 6.

16. The DNA molecule of claim 15 wherein said threonine occurring at the position corresponding to amino acid 249 of SEQ ID NO. 2 is replaced with an amino acid selected from the group consisting of isoleucine and alanine.

17. The DNA molecule of claim 11 wherein said second amino acid substitution occurs at the position corresponding to the proline at amino acid 118 of SEQ ID NO. 2 and said first amino acid substitution occurs at a position selected from the group consisting of

(a) the position corresponding to the alanine at amino acid 166 of SEQ ID No. 6; and

(b) the position corresponding to the tyrosine at amino acid 372 of SEQ ID No. 6.

18. The DNA molecule of claim 17 wherein said proline occurring at the position corresponding to amino acid 118 of SEQ ID NO. 2 is replaced with a leucine.

19. The DNA molecule of claim 11 wherein said second amino acid substitution occurs at the position corresponding to the asparagine at amino acid 425 of SEQ ID NO. 2 and said first amino acid substitution occurs at a position selected from the group consisting of

(a) the position corresponding to the alanine at amino acid 166 of SEQ ID No. 6; and

(b) the position corresponding to the tyrosine at amino acid 372 of SEQ ID No. 6.

20. The DNA molecule of claim 19 wherein said asparagine occurring at the position corresponding to amino acid 425 of SEQ ID NO. 2 is replaced with a serine.

21. The DNA molecule of claim 11 wherein said second amino acid substitution occurs the position corresponding to the tyrosine at amino acid 498 of SEQ ID NO. 2 and said first amino acid substitution occurs at a position selected from the group consisting of

(a) the position corresponding to the alanine at amino acid 166 of SEQ ID No. 6; and

(b) the position corresponding to the tyrosine at amino acid 372 of SEQ ID No. 6.

22. The DNA molecule of claim 21 wherein said tyrosine occurring at the position corresponding to amino acid 498 of SEQ ID NO. 2 is replaced with a cysteine.

23. The DNA molecule of any of claims 13-22 wherein said tyrosine occurring at the position corresponding to amino acid 372 of SEQ ID No. 6 is replaced with an amino acid selected from the group consisting of cysteine, isoleucine, leucine, threonine and methionine.

24. The DNA molecule of claim 12 wherein said alanine occurring at the position corresponding to residue 166 of SEQ ID No. 6 is replaced with an amino acid selected from the group consisting of valine, threonine, leucine, cysteine and tyrosine.

25. The DNA molecule of claim 12 wherein said glycine occurring at the position corresponding to residue 167 of SEQ ID No. 6 is replaced with a serine.

26. The DNA molecule of claim 12 wherein said glycine occurring at the position corresponding to residue 167 of SEQ ID No. 6 is replaced with a serine.

27. The DNA molecule of claim 12 wherein said cysteine occurring at the position corresponding to residue 161 of SEQ ID No. 6 is replaced with a phenylalanine

28. The DNA molecule of claim 12 wherein said isoleucine occurring at the position corresponding to residue 421 of SEQ ID No. 6 is replaced with a threonine.

29. The DNA molecule of claim 10 wherein said plant is selected from the group consisting of maize, wheat, soybean and *Arabidopsis*.

30 The DNA molecule of claim 10, wherein said plant protox comprises an amino acid sequence selected from the group consisting of SEQ ID Nos. 2, 4, 6, 8, 10 and 12.

31. A chimeric gene comprising a promoter active in a plant operably linked to a heterologous DNA molecule encoding a protoporphyrinogen oxidase (protox) selected from the group consisting of a wheat protox comprising the sequence set forth in SEQ ID No. 10 and a soybean protox, comprising the sequence set forth in SEQ ID No. 12.

32. The chimeric gene of claim 31 additionally comprising a signal sequence operably linked to said DNA molecule, wherein said signal sequence is capable of targeting the protein encoded by said DNA molecule into the chloroplast.

33. The chimeric gene of claim 31 additionally comprising a signal sequence operably linked to said DNA molecule, wherein said signal sequence is capable of targeting the protein encoded by said DNA molecule into the mitochondria.

34. A chimeric gene comprising a promoter which is active in a plant operably linked to the DNA molecule of claim 10.

35. The chimeric gene of claim 34 additionally comprising a signal sequence operably linked to said DNA molecule, wherein said signal sequence is capable of targeting the protein encoded by said DNA molecule into the chloroplast.

36. The chimeric gene of claim 34 additionally comprising a signal sequence operably linked to said DNA molecule, wherein said signal sequence is capable of targeting the protein encoded by said DNA molecule into the mitochondria.

37. A recombinant vector comprising the chimeric gene of claim 31, wherein said vector is capable of being stably transformed into a host cell.

38. A recombinant vector comprising the chimeric gene of claim 34, wherein said vector is capable of being stably transformed into a plant cell.

39. A host cell stably transformed with the vector of claim 37, wherein said host cell is capable of expressing said DNA molecule.

40. A host cell of claim 39 selected from the group consisting of a plant cell, a bacterial cell, a yeast cell, and an insect cell.

41. A plant comprising the DNA molecule of claim 10, wherein said DNA molecule is expressed in said plant and confers upon said plant tolerance to a herbicide in amounts which inhibit naturally occurring protox activity.

42. The plant of claim 41 wherein said DNA molecule replaces a corresponding naturally occurring protox coding sequence.

43. A plant comprising the chimeric gene of claim 34, wherein said chimeric gene confers upon said plant tolerance to a herbicide in amounts which inhibit naturally occurring protox activity.

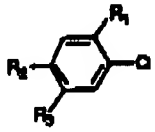
44. The plant of claim 41, wherein said plant is selected from the group consisting of maize, wheat, sorghum, rye, oats, turf grass, rice, soybean, cotton, tobacco, sugar beet, and oilseed rape.

45. A method for controlling the growth of undesired vegetation which comprises applying to a population of the plant of claim 41 an effective amount of a protox-inhibiting herbicide.

46. The method of claim 45 wherein said plant is selected from the group consisting of soybean, cotton, tobacco, sugar beet, oilseed rape, maize, wheat, sorghum, rye, oats, turf grasses and rice.

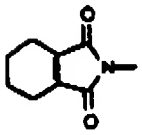
47. The method of claim 46 wherein said protox-inhibiting herbicide is selected from the group consisting of an aryluracil, a diphenylether, an oxidiazole, an imide, a phenyl pyrazole, a pyridine derivative, a 3-substituted-2-aryl-4,5,6,7-tetrahydroindazole, a phenopylate and *O*-phenylpyrrolidino- and piperidinocarbamate analogs of said phenopylate.

48. The method of claim 47 wherein said protox-inhibiting herbicide is an imide having the formula



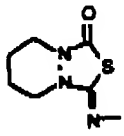
(Formula V)

wherein Q equals



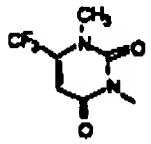
(Formula VI)

OR



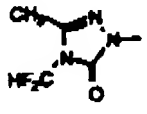
(Formula VII)

OR

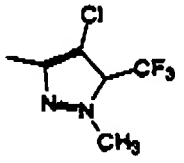


(Formula VIII)

OR

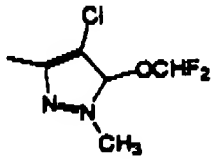


(Formula IX)



(Formula IXa)

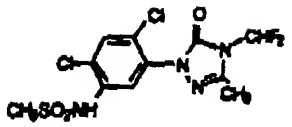
OR



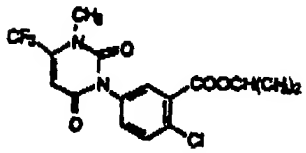
(Formula IXb)

and wherein R_1 equals H, Cl or F, R_2 equals Cl and R_3 is an optimally substituted ether, thioether, ester, amino or alkyl group, and wherein R_2 and R_3 together may form a 5 or 6 membered heterocyclic ring.

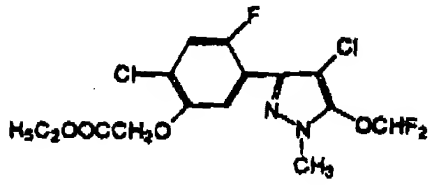
49. The method of claim 48 wherein said imide is selected from the group consisting of



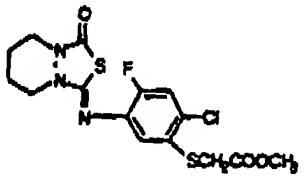
(Formula X);



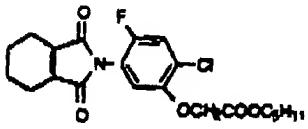
(Formula XI);



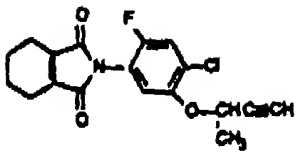
(Formula XII);



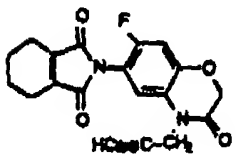
(Formula XIII);



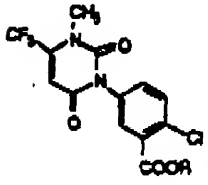
(Formula XIV);



(Formula XV);



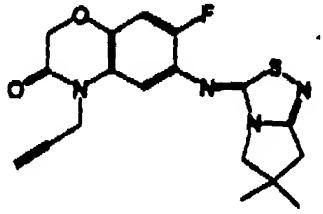
(Formula XVI); and



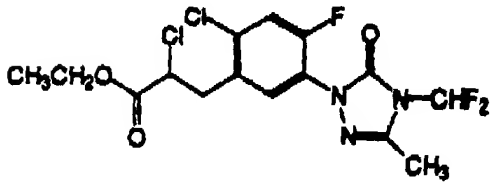
(Formula XVII)

wherein R signifies (C_{1-4}-alkenyloxy)carbonyl-C_{1-4}-alkyl.

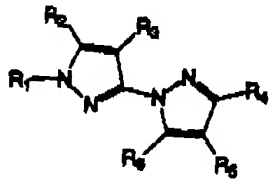
50. The method of claim 45 wherein said protox-inhibiting herbicide has the formula selected from the group consisting of



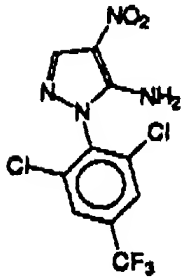
(Formula XVIII),



(Formula XIX),



(Formula XX), and



(Formula XXI).

ABSTRACT

The present invention provides novel DNA sequences coding for plant protoporphyrinogen oxidase (protox) enzymes from soybean and wheat. In addition, the present invention teaches modified forms of the protox enzyme which are herbicide tolerant. Plants expressing the herbicide tolerant protox enzymes taught herein are also provided. These plants may be engineered for resistance to protox inhibitors *via* mutation of the native protox gene to a resistant form or they may be transformed with a gene encoding an inhibitor-resistant from of the plant protox enzyme.